

ANALYSIS AND DESIGN OF CMOS
RADIO-FREQUENCY POWER AMPLIFIERS

A Dissertation

by

HAOYU QIAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jose Silva-Martinez
Committee Members,	Aydin Karsilayan
	Peng Li
	Duncan M. Walker
Head of Department,	Miroslav M. Begovic

May 2017

Major Subject: Electrical Engineering

Copyright 2017 Haoyu Qian

ABSTRACT

The continuous advancement of semiconductor technologies, especially CMOS technology, has enabled exponential growth of the wireless communication industry. This explosive growth in turn has completely changed people's lives. The CMOS feature size scale down greatly benefits digital logic integrations, which result in more powerful, versatile, and economical digital signal processing. Further research and development has pushed analog, mixed-signal, and even radio-frequency (RF) circuit blocks to be implemented and integrated in CMOS.

Future generations of wireless communication call for even further level of integration, and as of now, the only circuit block that is rarely integrated in CMOS along with other parts of the system is the power amplifier (PA). Due to the fact that the PA in a wireless communication system is the most power-hungry circuit block, the integration of RF PA in CMOS would potentially not only save the cost of the wireless communication system real estate, but also reduce power consumption since die-to-die connection loss can be eliminated.

RF PA design involves handling large amounts of voltage and current at the radio frequencies, which in the present wireless communication standards are in the range of giga-hertz. Therefore, a good understanding of many aspects related to RF PA design is necessary. Theoretical analysis of the communication system, nonlinear effects of the PA, as well as the impedance matching network is systematically presented. The analysis of the nonlinear effects proposes a formal mathematical description of the multitone nonlinearity, and through its relationship with two-tone test, the proposed PA design methodology would greatly reduce the design time while improving the design accuracy.

A thorough analysis of the available architecture and design techniques for efficiency and linearity enhancement of RF PA shows that despite tremendous amounts of research and development into this topic, the fundamental tradeoff between the two still limits the RF PA implementation largely within SiGe, GaAs, and InP technologies. A RF PA for Wideband Code-Division Multiple Access (WCDMA) application standard is proposed, designed, and implemented in CMOS that demonstrates the proposed segmentation technique that resolved the main tradeoff between power efficiency and linearity. The innovative architecture developed in this work is not limited to applications in the WCDMA communication protocol or the CMOS technology, although CMOS implementation would take advantage of the readily available digital resources.

To my wife, parents, and late grandparents.

ACKNOWLEDGMENTS

First, I would like to acknowledge my advisor, Professor Jose Silva-Martinez. It was Dr. Silva who first introduced me to analog integrated circuit design, and through my entire PhD program, I had always looked up to his passion, his approach to solve problems, and his dedication. I am very grateful to have had a chance to work with him, and am thankful for his guidance that made this research possible.

I would like to thank Professor Aydin Ilker Karsilayan for his valuable suggestions and technical insights. I also greatly appreciate Professor Peng Li and Professor Duncan M. Walker for their comments and inputs, and for taking time out of their very busy schedules to serve on my graduate committee.

I feel very honored to be part of the AMSC group, and want to thank Heng Zhang, Yung-Chung Lo, Shan Huang, Jiayi Jin, Cheng Li, Chengliang Qian, Jie Zou, Xi Chen, Zhuizhuan Yu, Yang Liu, Jingjing Yu, Hongbo Chen, Jun Zhou, Yang Su, Chen Ma, Jun Yan, Miao Song, Hao Huang, Yang Gao, Yanjie Sun, Nan Wang, Geng Tang, Congying Shi, Xiaosen Liu, Younghoon Song, John Mincey, Marvin Onabajo, Chang-Joon Park, Sai Ganta, Jesse Coulon, Mohan Geddada, Alex Edward, Carlos Briseno, Ehsan Tabasy, Nagar Rashidi, Hajir Hedayati, Efrain Gaxiola, Jorge Zarate, and Richard Turkson for their friendship. In particular, I want to thank Jingjing Yu, Chen Ma, Jun Yan, and Jesse Coulon for many nights spent together on course work. Also I want to thank Shan Huang for many invaluable technical discussions. Younghoon Song and John Mincey also served as my TA during my graduate study, and with their countless help to the very details of many aspects of analog IC design, I gained the most out of the classes. I had the honor of working with Dr. Marvin Onabajo on his research project of RF mixer built-in self testing. He is the most hard-working person I have ever worked with, and I learned quite a lot from him. Also, I really want to thank Professor Kamran Entesari for taking time to share his expertise in impedance matching techniques. In addition, I want to express my appreciation of Ella Gallagher for facilitating events and paper work.

I would like to express my gratitude to the Department of Electrical & Computer Engineering of Texas A&M University for providing such a good environment for academic research, and for offering such a comprehensive course work that prepares me for my career. In particular, I would like to thank Professor Scott Miller for teaching me Digital Communication Theory, and Professor Aniruddha Datta for teaching me Control Theory. The fundamental concepts in these classes greatly helped me with my graduate research. I am very grateful to Tammy Carda to her assistance all the way through my graduate study. Her administrative competence and warm personality greatly eased my life as a graduate student.

During my internship at Microtune, Jason Wardlaw was my mentor, and had helped me quite a lot to get to speed. Also I want to thank Yan Cui, Jan-Michael Stevenson, and Ron Spencer for technical discussions, and Kirk Asby for his great leadership of the many good people at Microtune and for giving me the opportunity to work with them.

Outside of the Department of Electrical & Computer Engineering, I would like to thank Professor Joseph H. Ross, Jr. from Department of Physics. Dr. Ross was my research advisor during my study at Department of Physics, and it was him who first taught me how research should be done at the world-class level. I really appreciate his mentorship and unselfishly encouragement to pursue the Electrical Engineering degree.

Outside of Texas A&M University, I want to express my gratitude toward Eric Soenen of TSMC for giving me the opportunity to tape out my project with their 40 nm CMOS technology. Also I want to thank Sherif Embabi of Nvidia for allowing me to measure the power amplifier at their lab with their instruments. Their help all directly contributed to my work.

Finally, I must reserve the most special appreciation for my friends and family. I thank my friends at Texas A&M University and elsewhere for their friendship. I am deeply indebted to my parents, Guyuan Lu and Cheng Qian, who gave me a loving family and have always emphasized the importance of education. My late grandparents, Maozhen Zhang and Shaojie Lu, brought me up, and had supported me every step of the way. At last, my most tender and sincere thanks are reserved for my wife, Sulei Chen, who was willing to marry a graduate student, but nevertheless continues giving me her unconditional love and support.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Jose Silva-Martinez, Aydin Karsilayan, and Peng Li of the Department of Electrical and Computer Engineering, and Professor Duncan M. Walker of the Department of Computer Science and Engineering.

All work for the dissertation was completed by the student, under the advisement of Professor Jose Silva-Martinez.

Funding Sources

Graduate study was supported by teaching assistantship from Texas A&M University.

TABLE OF CONTENTS

	Page
I INTRODUCTION	1
I.A Background	1
I.B Motivation and Challenges in CMOS Power Amplifier (PA)	4
I.C Orgnization of the Dissertation	5
II PA CLASSIFICATIONS	8
II.A Introduction	8
II.B Class A Power Amplifier	9
II.C Class B Power Amplifier	10
II.D Reduced Conduction Angle PAs	13
II.E Class D Power Amplifier	14
II.F Class E Power Amplifier	16
II.G Class F Power Amplifier	17
III DIGITAL COMMUNICATION SYSTEMS	20
III.A Introduction	20
III.B Digital Modulation	20
III.C Wireless Communication Standards	24
IV NONLINEAR EFFECTS OF POWER AMPLIFIER	27
IV.A Introduction	27
IV.B MOSFET Nonlinearity Mechanisms	28
IV.C Single-tone Test: Harmonic Distortion and Gain Compression	30
IV.D Two-tone Test: Intermodulations and Intercept Point	33
IV.E Multitone Nonlinear Effects	36
IV.F AM-to-PM Conversion	46
V IMPEDANCE MATCHING NETWORK DESIGN	48
V.A Introduction	48
V.B Impedance Matching Theory	48
V.C L-match Network	49
V.D Pi-match Network	53
V.E Multisection Matching Network	59
VI POWER AMPLIFIER ARCHITECTURES	61
VI.A Introduction	61
VI.B Efficiency Enhancement Techniques	61
VI.C Linearization Techniques	69

VII	A 35DBM OUTPUT POWER AND 38DB LINEAR GAIN PA WITH 44.9% PEAK PAE AT 1.9GHZ IN 40 NM CMOS	72
VII.A	Introduction	72
VII.B	Efficiency Enhancement Techniques	73
VII.C	PA Architecture	75
VII.D	PA System Design	82
VII.E	Measurement Results	88
VII.F	Conclusion	93
VIII	CONCLUSION	95
	REFERENCES	97

LIST OF FIGURES

FIGURE	Page
1 Global cellular subscription growth.	2
2 Distribution of global mobile subscribers by technology.	2
3 Global smartphone shipments forecast.	3
4 Schematic of a generic single-ended PA.	8
5 Class A drain current and voltage waveforms at maximum output level.	10
6 Class B current and voltage waveforms at maximum output level.	11
7 Efficiency of classes A and B as function of PBO.	12
8 Reduced conduction angle PA current and voltage waveforms at maximum output level.	13
9 Maximum efficiency and output power as a function of conduction angle.	15
10 Conceptual schematic of a class D PA.	15
11 Class D current and voltage waveforms.	16
12 Conceptual schematic of a class E PA.	17
13 Class E current and voltage waveforms.	17
14 Class F current and voltage waveform illustrations.	19
15 Block diagram of a digital RF transmitter.	20
16 Relationship between two representations of complex envelope.	21
17 Ideal BPSK constellation diagram.	23
18 Ideal QPSK constellation diagram.	23
19 Ideal 16QAM constellation diagram.	24
20 Illustrative I-V characteristic of a MOSFET.	28
21 Typical power transfer characteristics of a PA.	32
22 Third-order intermodulation.	34
23 Geometrical interpretation of IP_3 calculation.	35
24 Typical RF transmitter.	36
25 Microphotograph of the chip.	44
26 Simulated and predicted multitone ACLR as a function of number of tones.	44
27 Multitone output spectrum.	45
28 PA output impedance matching block diagram.	49

29	Lumped-element L-match network.	50
30	Concept of loaded Q	51
31	Lumped L-match driven by the intended source resistance.	51
32	Hybrid L-match network.	52
33	Π -match network in output matching applications.	53
34	General schematic for analysis of a Π -match network.	54
35	Π -match network with inductor's parasitic resistance explicitly shown.	58
36	Thevenin equivalent circuit of the Π -match network near matching frequency.	58
37	Multisection matching network.	59
38	Schematic of a 2-stage matching network.	60
39	Conceptual schematic of outphasing modulation.	62
40	Demonstration of outphasing load configuration without compensation.	63
41	Outphasing load compensation network.	63
42	Outphasing technique implemented in the current domain.	64
43	Conceptual schematic of a Doherty PA.	65
44	Simplified Doherty PA for analysis.	65
45	Conceptual schematic of an EER architecture.	67
46	Conceptual schematic of a polar modulation architecture.	68
47	Conceptual schematic of an ET system.	68
48	Conceptual schematic of a power combining architecture.	69
49	Conceptual schematic of a polar-loop feedback system.	69
50	Conceptual schematic of a cartesian feedback system.	70
51	Conceptual schematic of a feedforward architecture.	71
52	Conceptual schematic of a predistortion system.	71
53	Conceptual schematic of an ET system.	74
54	Conceptual schematic of power combining architecture with switchable PAs.	75
55	Correlation between control phases and baseband signal amplitude.	76
56	Simplified schematic of the proposed architecture.	77
57	Simplified model for timing mismatch analysis.	78
58	PA output waveforms (RF component is not shown for simplicity).	79
59	Timing mismatch effects on ACLR.	81

60	Schematic of the PA output stage; the core consists of 1536 replicas.	82
61	Schematic showing the CMFB circuit allocated at PA output.	83
62	Conceptual schematic of the driver stage.	84
63	Simulation results of common-mode voltage transient response.	85
64	Two-section impedance matching network.	86
65	Insertion loss simulation with process variations.	86
66	Transient simulation results.	87
67	Microphotograph of the chip.	88
68	Measured gain, output power, and PAE as a function of input at 1.9 GHz.	88
69	Simulation and measurement results of PA's S22.	89
70	ACLR measured at maximum output power of 31 dBm.	90
71	SEM measured at maximum output power of 31 dBm.	90
72	ACLR as a function of maximum output power.	91
73	EVM as a function of maximum output power.	92
74	Phase error as a function of maximum output power.	93

LIST OF TABLES

TABLE	Page
I RF/Microwave-Related Publications in IEEE Database	3
II QPSK Format Summary	23
III GSM PA Specifications	25
IV IEEE 802.11g PA Specifications	26
V WCDMA PA Specifications	26
VI Intermodulation Adjacent Tone Position Analysis	40
VII Triple-beat Adjacent Tone Position Analysis	41
VIII Impedance Matching Network Component Values	86
IX Comparison With Recently Published Works	93

I. INTRODUCTION

I.A. Background

Radio frequency (RF) circuits and systems have been around for more than a century. Since the theoretical prediction by Maxwell and later experimental verification by Hertz of the electromagnetic waves, scientists and engineers have put endless endeavor to develop systems that are able to transmit and receive information embedded in such waves. From the inventions of electrical telegraph to radio broadcasting, then telephone and television broadcasting, internet, cellular phone, and nowadays global positioning system (GPS), bluetooth, and wireless local area network (WLAN), the list that shows the effort to enable and improve means of communication via RF and microwave technologies keeps growing.

The first-generation (1G) cellular system first appeared in 1969 [1]. Then after a decade of patenting, licensing, and trial services, the first full-service 1G cellular system finally took place in late 1983 [2]. The following generations were developed and put to market in an exponentially faster pace. Development in semiconductor technologies is a deciding factor: ever since the invention of integrated circuit (IC), especially the development of the CMOS IC, digital signal processing (DSP) saw great advance and thus more functionalities can be integrated. Mobile device and hence services became more affordable. As a result, today's mobile devices are no longer limited in making telephone calls, but can have functionalities such as bluetooth, WLAN, GPS, and so on. On the other hand, the frequency band originally intended only for handling voice information can realize video calls now, and at the same time photos, videos, and files can be shared. Fig. 1 shows the growth of cellular subscriptions over the last decade. Not only has the total number of cellular subscriptions increased by a factor of about seven during the past 14-year span, the subscriptions per 100 inhabitants is approaching 100. Note that the cellular industry did not show significant slow-down even during the global economic recession in the late 2000s. The size of the population does not pose as a limit to the expansion of market either. Fig. 1 shows that the *average* subscription percentage is close to 100% as projected in the year 2014, and detailed statistics show that in some countries and regions this number has already exceeded the 100 mark for several years.

With the development of the third-generation (3G) wireless communication standards, wireless communication has become so reliable, versatile, and thus convenient, that it is no longer considered a luxury or even optional, but has become an essential part of people's everyday lives. With the introduction of the fourth-generation (4G) technology that applies not only to mobile phones (especially smartphones) but also other devices such as tablet, laptop, television, and even motor vehicle, the concept of wireless

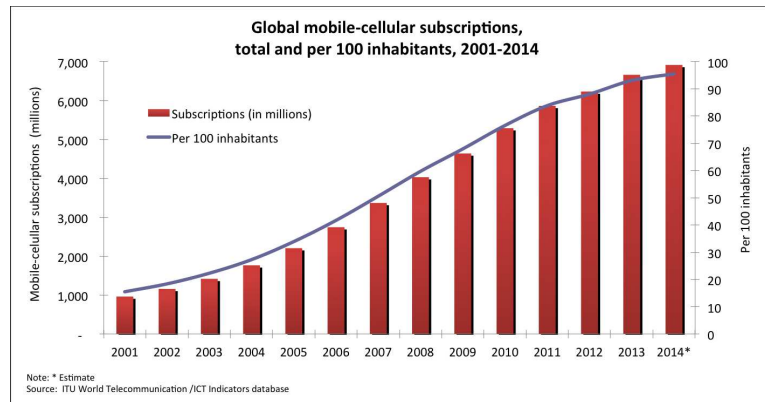


Fig. 1. Global cellular subscription growth. [3]

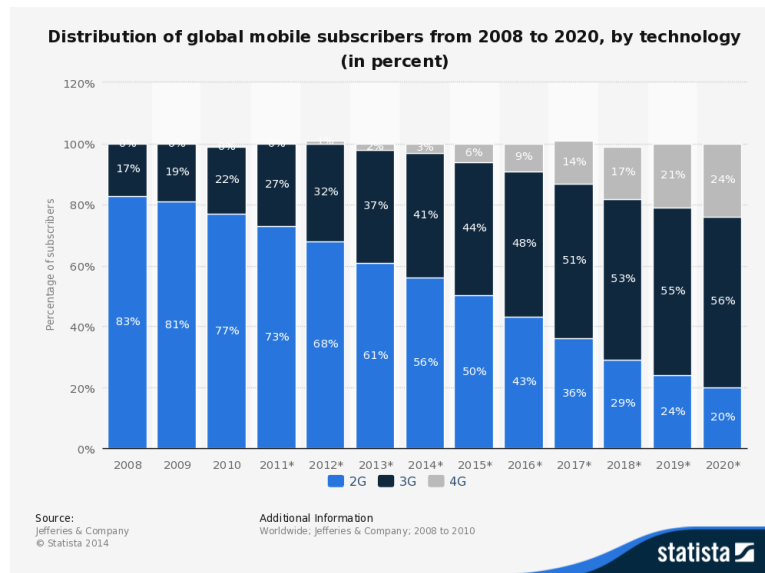


Fig. 2. Distribution of global mobile subscribers by technology. [4]

communication is extended to communication between devices. As predicted in Fig. 2, new generations of wireless communications standards steadily take over the old ones. The fast pace of technology take-over in the wireless communication has uniquely made it one of the most active, dynamic, and exciting in the semiconductor industry.

The promising functionalities in the 3G and 4G technologies indeed imply a promising potential in the wireless market. Fig. 3 shows the global smartphone shipments forecast. Despite the mobile market in all developed countries and most developing countries being already mature, the smartphone shipments still see a steady increase in the forecast.

The growth in market and continuous evolution of wireless communication technologies in turn pushed

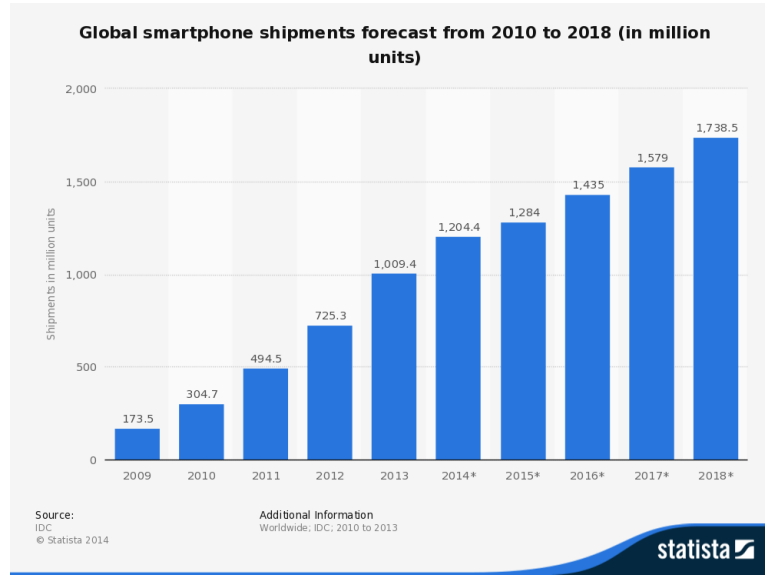


Fig. 3. Global smartphone shipments forecast. [5]

for more research and development in RF and microwave IC and system design. Table I shows the resultant number of publications of a simple search of the keywords “radio frequency” and “microwave” in the IEEE Xplore database. In these fields, the number of IEEE publications in the 1990s decade is more than 1.5 times the total number before 1990, the millennium decade almost tripled as compared to the 1990s decade, and the number of IEEE publications within less than the recent four years has already surpassed the entire 1990s decade.

TABLE I
RF/Microwave-Related Publications in IEEE Database

Year	No. of publications
Before 1990	2953
1991 to 2000	4498
2001 to 2010	12099
2011 to present	5095

One important factor that has been a constant driving force of the aforementioned growth in both the industry and research and development effort is the advance in semiconductor technology. Because of the scaling down of the feature size, more transistors can be placed per unit area, and the overall cost is reduced. The major beneficiary of such progresses is integrated digital circuits because more gates and thus functionalities can be integrated. As a result, the DSP has become more powerful, versatile, and economical. For analog and RF front-end designs, the scaling of semiconductor presents more of a design trade-off than one-sided benefit their digital counterpart experiences. Therefore the front-end and

back-end used to be implemented on separate chips, each using a semiconductor technology optimized for their respective performance metrics. Although such a degree of integration greatly reduced cost as compared with discrete circuits, it will not be able to keep pace with the growing market demand for even lower cost to have enough profit margin. Consequently, the trend has been towards integration of digital and analog even RF circuitries on the same die, termed *system on chip* (SoC). Analog and mixed signal circuits are first integrated with the digital back-end, then small-signal RF circuitry, such as the RF receiver [6–10] and transmitter driver [11–15], followed suit. In addition to savings in power consumption and area, RF circuits fits naturally to today’s high-speed digital because they can share at least part of the clocking circuits, and the increasingly more common RF-to-digital interface and digital assistance to the RF circuits such as calibration, predistortion, built-in self testing, and multimode controls would be bulky and inefficient if the RF and digital sections are on separate dice [16].

Since most of the signal processing is performed in the digital domain, and the CMOS technology is best suited for digital ICs to this date, the aforementioned trend is then translated into the integration of the system on digital CMOS die. Analog and mixed-signal building blocks such as amplifier, active filter, bandgap reference, voltage regulator, and data converter have been successfully implemented in digital CMOS with satisfactory performance and low power consumption. Due to lossy and thus noisy substrate and relatively large and nonlinear parasitic component as compared with conventional RF IC technologies such as SiGe, GaAs, and InP, CMOS RF IC had not been developed and put to industry at the early stage of this process of integration. But lower cost as a result of integration and better battery life as a consequence of low supply voltage and bias current of the CMOS technology fit the demand of the mobile communication market so well that a great amount of effort of research and development has resulted in reliable and high-performance small-signal RF IC implemented in CMOS.

I.B. Motivation and Challenges in CMOS Power Amplifier (PA)

RF PAs are the biggest power consumer in the RF transceiver chain and occupy large die area, but they have large current and voltage swing at high frequency, hence are difficult to implement in digital CMOS. Although there is an increasing interest and research effort in the academic community, commercial PAs are still dominated by SiGe and GaAs technologies as of today. The drawback of CMOS technology rooted in the reasons mentioned above is a factor, and another reason that is holding off commercial CMOS PAs and thus full integration of the entire radio is the reliability issue due to CMOS scaling. As the CMOS technology scales down to finer minimum dimensions, the maximum allowable voltage

is lowered, and various mechanisms have made CMOS PAs especially prone to device breakdown [17]. Consequently, realizing high-power CMOS PAs with conventional architectures would result in very low load impedance, which is difficult to realize, prone to process, voltage, and temperature (PVT) variations, and can have high power loss. In addition, 3G and 4G standards use modulation schemes that have high peak-to-average power ratio (PAPR), thus highly linear PAs are needed. Multimode requirements in the future generations of mobile devices would add flexibility requirements to the PA design, making it more challenging to be implemented in digital CMOS. Since the majority of the CMOS foundry's customers use the technology for digital or mixed-signal applications, the modeling and characterization from the foundry are also based on those applications, which is not enough even for general RF IC applications, not to mention high-power RF PAs. Therefore another challenge in CMOS PA design is the lack of accurate modeling and thus simulation tools. For example, even in a CMOS technology that comes with S-parameters, they are small-signal S-parameters, which can only serve as a reference for PA design. GaAs and SiGe technologies suffer less from such a problem because they are mainly used for RF PA applications and their modeling emphasizes on large-signal RF behavior.

Despite the challenges discussed above, because of the potential cost-reduction in CMOS integration, there is a lot of research effort put into the design of CMOS PAs. Along the way, various techniques have been developed to not only alleviate the limitations posed by the CMOS technology but also improved PA performance. Therefore, in addition to finding a low-cost solution to fully integrated wireless transceiver, another motivation of this work is that the techniques developed in improving PA performance metrics such as linearity and efficiency in CMOS technology, which is not optimized for RF PA applications, may also find their enduring values in other semiconductor technologies, future CMOS nodes, or future semiconductor technologies, and even in other applications.

Some call CMOS PAs the “last frontier”, or the “missing piece” of full integration of the wireless transceiver system. This dissertation hence presents the analysis of various aspects of RF PA design, and introduces the research results of a segmented CMOS PA for wireless communication applications.

I.C. Organization of the Dissertation

The objective of this work is to exploit standard CMOS technology for development of RF PAs for current and future wireless communication applications. RF PA design is not an easy task because the designer is required to have a solid understanding of analog, digital, and RF IC design concepts. In addition, due to the challenges in realizing CMOS RF PAs mentioned above, it is important to be aware

of the device physics and various failure mechanisms in modern CMOS technology. Moreover, proficiency is needed in other areas such as microwave, communication, and signal processing, each of which is itself a well-developed discipline. This dissertation would not cover all topics related to CMOS RF PA design in great details, but would present an analysis of some of the important ones before discussing the research results of a highly linear and efficient CMOS PA designed for high-power wireless communication applications.

Section II first introduces various operation modes of PAs, which serve as a starting point of the discussion and analysis of RF PA design. Each operation mode has its own advantages and disadvantages, and the discussion of all those operation modes would reveal that important specifications of RF PA design usually trade off with each other.

Section III provides an introduction to digital communication systems. Modern communication systems are mostly implemented using digital modulation schemes, and modulation theory is a well-developed discipline that deserves a discussion in the length of a textbook [18]. This section only lays out some of the most critical modulation concepts and most common modulation schemes used in wireless communications, and how they relate to RF PA design specifications. One of the requirements of modern communication standards revealed in this section is the stringent specification of linearity, thus a discussion of nonlinear effects of PAs is presented in Section IV.

The root-cause of nonlinear effects lies in the use of nonlinear active device. Therefore the nonlinear mechanisms of MOS transistors is first discussed. Conventional linearity tests, i.e. single-tone and two-tone tests are also briefly reviewed. As would be mentioned in Section III, modern communications use wideband, multicarrier modulation schemes, hence the conventional linearity tests can only provide an indication of the PA linearity. On the other hand, a multitone test or modulated envelope simulation require long simulation time and large computation resource. To resolve this issue, a detailed analysis of multitone nonlinear effects is carried out in this section, and as the analysis shows, a simple two-tone test can be used to predict multitone nonlinear behavior. Although currently not a major nonlinearity contributor to CMOS PA, the amplitude modulation to phase modulation (AM-to-PM) conversion is briefly analyzed at the end of this section.

Section V presents an analysis of PA's output impedance matching. The output impedance matching is especially important for PA because it is usually the last circuit network before T/R switch or the antenna, and its insertion loss hurts the power efficiency the most. As mentioned previously, CMOS PA's optimal output impedance at the drain is usually very small due to low-voltage technology. Therefore a

high impedance transformation ratio results, and thus the quality factor Q of the impedance matching network is high. High- Q networks have narrow bandwidth, and such a network may be sensitive to PVT variations. To alleviate this issue, a multisection impedance matching network is proposed, and simulation results show a wide bandwidth and insensitivity to PVT variations.

Because the sophisticated multifunctional wireless electronic devices demand high efficiency PA for longer battery life, and today's strict wireless communication standards and overly crowded communication frequency bands require highly linear transceivers, conventional standalone PAs would not meet all the requirements. Various PA architectures aiming at improving power efficiency and linearity have been developed, and Section VI reviews these techniques and their advantages and disadvantages.

Based on all the analysis and discussion of all the important issues concerning CMOS RF PA design, Section VII presents a research project that develops a solution that improves the PA's power efficiency while keeps the linearity performance. PA segmentation technique is proposed, whose control scheme is correlated with the input signal power level. A fast switching scheme ensures the PA segments can be activated and deactivated within the modern wideband wireless communication standards, therefore the average power efficiency can be improved both within a specific standard or in a multimode application. Such a proposal and solution is implemented in TSMC 40 nm CMOS technology and supported by good silicon measurement results of a WCDMA signal.

Finally, Section VIII summarizes and concludes this dissertation.

II. PA CLASSIFICATIONS

II.A. Introduction

Fig. 4 shows the schematic of a generic single-ended PA. The active device can be implemented using MESFET, HEMT, pHEMT, BJT, JFET, or MOSFET, but since the focus of this dissertation is on CMOS PA design, an NMOS transistor symbol is shown in the figure. The choke inductor (RFC) and blocking capacitor (C_B) ensure isolation of DC and RF signal paths. The output filter usually has multiple functions: a) it attenuates out-of-band components; b) it enables impedance transformation, so that the impedance of the antenna Z_L can be transformed to the optimal output impedance the transistor sees at the drain or collector, Z_T ; and c) it realizes infinite or zero impedance at multiples of the harmonics of the RF signal, so that the output waveform can be shaped.

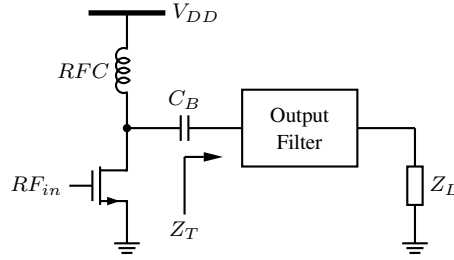


Fig. 4. Schematic of a generic single-ended PA.

Based on the method of operation, conventional RF PAs are categorized into classes A - F [2, 19, 20]. In classes A, AB, B, and C PAs, the transistor operates as a current source, whereas in classes D, E, and F, the transistor is utilized as a switch, and therefore those PAs are generally termed switching mode PAs. Current-source type PAs, especially classes A and B, are inherently linear, but their power efficiencies degrades in the power back-off region. On the other hand, switching mode PAs usually can achieve power efficiencies that are theoretically close to 100%, but do not preserve amplitude linearity.

The main point of comparison between various classes of PAs is their drain efficiencies (DE), defined as the output power at the fundamental tone to the DC power:

$$\eta = \frac{P_{out}}{P_{DC}} \quad (1)$$

The power-added efficiency (PAE) is also a metric of the efficiency of the PA, defined as the ratio of the power difference at the fundamental frequency between the output and input of the PA to the DC power consumption:

$$PAE = \frac{P_{out} - P_{in}}{P_{DC}} = \eta \left(1 - \frac{1}{G_p} \right) \quad (2)$$

where G_p is the power gain of the PA. This definition of efficiency also includes the power gain considerations, and is approximately equal to η if G_p is large. Theoretical analyses in this section will compare the drain efficiencies of various classes of PA operation, because the power gain information, which is usually difficult to obtain in idealized analysis, is not needed, whereas in the design example discussed in a later section, PAE will be reported.

In modern wireless communication applications, the non-constant envelope signals are more likely to be applied to the PAs. Therefore, in addition to analyzing the maximum efficiency a certain class of PA can obtain, it is also important to investigate the power efficiency as a function of the output power back-off (PBO), defined as how much in decibel the output power is less than the maximum output power, or

$$PBO = 10 \log \frac{P_{out,max}}{P_{out}} \quad (3)$$

Since different modulation schemes would result in different peak-to-average power ratio ($PAPR$) of the envelope, the average efficiency as a function of $PAPR$ provides a hint to making design decisions towards a specific modulation scheme. Mathematically, the $PAPR$ can be expressed as

$$PAPR = 10 \log \frac{P_{out,max}}{P_{out,avg}} \quad (4)$$

Basic operations of various classes of PAs will be described in this section, while techniques to improve back-off region power efficiencies for current source PAs and to enhance linearity for switching mode PAs will be discussed in a later section.

II.B. Class A Power Amplifier

Class A PAs are biased such that the transistor is in the active region at all specified input levels. Assume the input voltage is sinusoidal, the ideal drain voltage and current at the maximum output level are shown in Fig. 5, where $\theta = \omega t$ for simplicity. Since the voltage across the choke inductor can be both positive and negative, the drain voltage can swing between 0 and $2V_{DD}$. Accordingly, the drain current varies between I_{max} and 0, with I_{max} depending on the bias and loading conditions as well as the transistor size.

From Fig. 5, the maximum efficiency of an ideal class A PA is achieved at the maximum output power:

$$\eta_{A,max} = \frac{\frac{1}{2}(I_{max}/2)V_{DD}}{(I_{max}/2)V_{DD}} = 50\% \quad (5)$$

The DC power consumption is fixed, so the efficiency as a function of PBO and the average efficiency

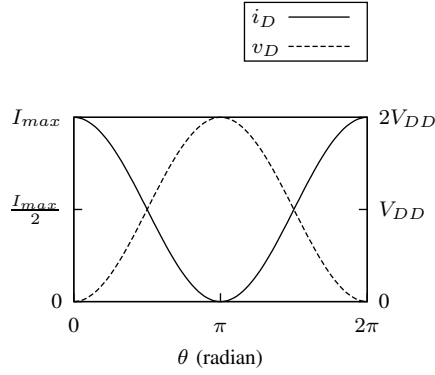


Fig. 5. Class A drain current and voltage waveforms at maximum output level.

as a function of PAPR are

$$\eta_A = \frac{P_{out}}{P_{DC}} = \frac{P_{out}}{P_{out,max}} \cdot \frac{P_{out,max}}{P_{DC}} = \eta_{A,max} 10^{-PBO/10} \quad (6)$$

$$\eta_{A,avg} = \frac{P_{out,avg}}{P_{DC}} = \frac{P_{out,avg}}{P_{out,max}} \cdot \frac{P_{out,max}}{P_{DC}} = \eta_{A,max} 10^{-PAPR/10} \quad (7)$$

So the efficiency of class A PAs would decay very quickly as the output power drops into the PBO region, for example, the ideal efficiency of a class A PA at 6 dB power back-off from the maximum output power would drop from 50% to 12.5%, only a quarter of the maximum efficiency. Also, class A average efficiency for high PAPR modulation schemes is very low. For instance, multi-channel OFDM modulations yield a PAPR of about 10 dB, so the ideal class A average efficiency would be only 5%, or a tenth of its maximum value. Note that the above analysis is based on the ideal class A operation, which ignored the finite minimum V_{DS} that is needed to keep the transistor in the active region. As the CMOS technology scales, $V_{DS,min}$ has become a fraction of V_{DD} that is no longer negligible, and $\eta_{A,max}$ would be less than 50%, resulting in even less average efficiency in high PAPR modulation schemes.

II.C. Class B Power Amplifier

If the PA only conducts half of the RF cycle it is of class B. The ideal drain current and voltage waveforms at maximum output level is shown in Fig. 6. Since the drain current is a half-wave, there will be harmonics. In class B PA analysis, it is assumed that all harmonics are short-circuited by the output filter. Therefore the drain voltage of a class B PA is still an ideal sinusoid.

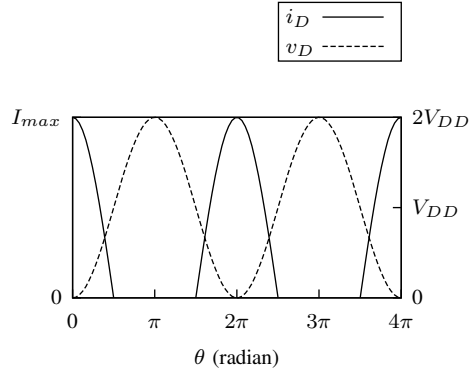


Fig. 6. Class B current and voltage waveforms at maximum output level.

The drain current is expressed as

$$i_{DS} = \begin{cases} I_{max} \cos \theta, & -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2} \\ 0, & -\pi \leq \theta < -\frac{\pi}{2}, \frac{\pi}{2} < \theta \leq \pi \end{cases} \quad (8)$$

where for convenience of the analysis, the $[-\pi, \pi]$ RF cycle is chosen. The DC and fundamental components of the drain current are obtained by calculating the corresponding Fourier coefficients:

$$I_{DC} = I_0 = \frac{1}{2\pi} \int_{-\pi/2}^{+\pi/2} I_{max} \cos \theta d\theta = \frac{I_{max}}{\pi} \quad (9a)$$

$$i_{RF} = I_1 = \frac{2}{2\pi} \int_{-\pi/2}^{+\pi/2} I_{max} \cos^2 \theta d\theta = \frac{I_{max}}{2} \quad (9b)$$

Therefore the theoretical maximum power efficiency of a class B PA is

$$\eta_{B,max} = \frac{\frac{1}{2} (I_{max}/2) V_{DD}}{(I_{max}/\pi) V_{DD}} = \frac{\pi}{4} \approx 78.5\% \quad (10)$$

But the advantage of class B over class A is not only its higher maximum efficiency. From the derivation of (9a), its DC current consumption is correlated with the output current. Suppose the peak output current in the PBO is I_{pk} , then

$$P_{DC} = \frac{I_{pk}}{\pi} V_{DD} = \frac{I_{pk}}{I_{max}} P_{DC,max} \quad (11)$$

The RF current at the fundamental is the same as (9b) except I_{max} is replaced I_{pk} , whereas the voltage amplitude would be the product of the RF current amplitude and the optimal load impedance. Therefore

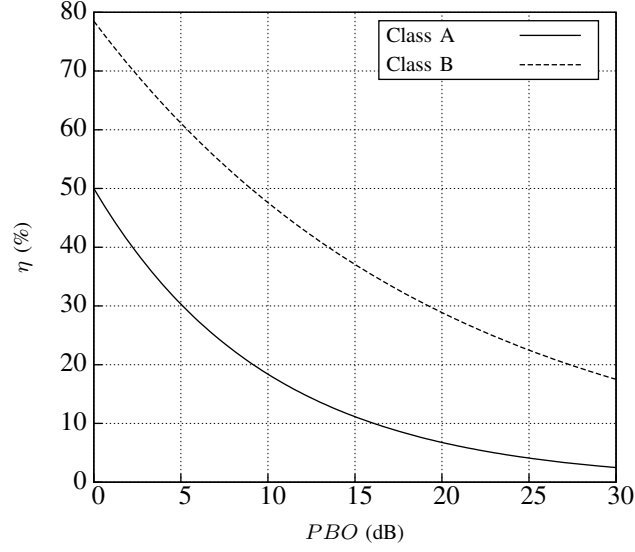


Fig. 7. Efficiency of classes A and B as function of PBO.

the output power can be expressed as

$$P_{out} = \frac{1}{2} \left(\frac{I_{pk}}{2} \right)^2 \frac{V_{DD}}{I_{max}/2} = \left(\frac{I_{pk}}{I_{max}} \right)^2 P_{out,max} \quad (12)$$

Combination of (11) and (12) would lead to the power efficiency of class B PAs as a function of PBO:

$$\eta_B = \eta_{B,max} \left(\frac{I_{pk}}{I_{max}} \right) = \eta_{B,max} \sqrt{\frac{P_{out}}{P_{out,max}}} = \eta_{B,max} 10^{-PBO/20} \quad (13)$$

Fig. 7 shows theoretical power efficiency of classes A and B as a function of PBO, according to (6) and (13). Class B has a larger maximum efficiency, and more importantly, it decays more slowly than class A as the PA enters PBO region. The class B average efficiency as a function of PAPR is

$$\eta_{B,avg} = \eta_{B,max} 10^{-PAPR/20} \quad (14)$$

One of the shortcomings of class B PA is the difficulties in reliable and insensitive realizations. As the CMOS technology scales, the transistor current shows more gradual variations around the threshold voltage, thus simply biasing the transistor gate at the threshold voltage would result in a drain current waveform that is far from the ideal case. Even the drain current leakage below or close to the threshold voltage is negligible, the threshold voltage itself varies by as much as 50% due to process, voltage, and

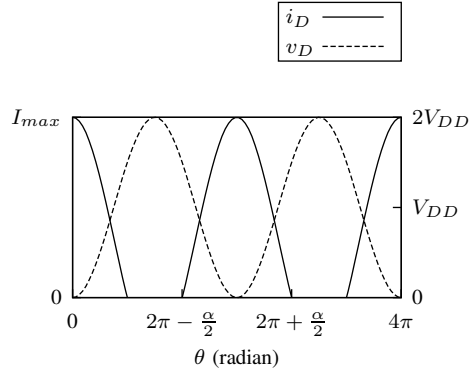


Fig. 8. Reduced conduction angle PA current and voltage waveforms at maximum output level.

temperature variations, hence a robust realization could be a challenge. Finally, the ideal class B amplifier is linear based on the assumption that harmonics are all short-circuited by the output filter. Such an assumption requires the output filter have relatively high quality factor. In low voltage applications, the optimal load resistance is low for high output power, thus the output filter would not have high Q , and thus the linearity of the class B PA may not be guaranteed.

Another disadvantage of class B as compared with class A is that to deliver the same RF power at the fundamental, class B PA needs twice the input voltage amplitude as required by class A. In other words, the power gain of class B PAs is 6 dB less than that of class A.

II.D. Reduced Conduction Angle PAs

The aforementioned classes A and B operations can be viewed as special cases of a more general concept. Define *conduction angle* α to be the proportion of the RF cycle for which conduction occurs [20], then the class A amplifiers are the ones with $\alpha = 2\pi$, whereas for class B, $\alpha = \pi$. In general, for a sinusoidal input, the current waveform is a truncated sinusoid if $\alpha < 2\pi$, as shown in Fig. 8. Not surprisingly class AB is defined as $\pi < \alpha < 2\pi$, while for PAs whose $0 < \alpha < \pi$ they are of class C. Since the transistor does not conduct all the time, the efficiency performance is expected to be better than that of class A. For reduced conduction angle PA analyses, it is assumed again that all harmonics are short-circuited by the output filter, therefore the voltage waveform in Fig. 8 is still an ideal sinusoid between 0 and $2V_{DD}$.

Mathematically, the drain current can be modeled as

$$i_{DS} = \begin{cases} I_Q + I_m \cos \theta, & -\frac{\alpha}{2} \leq \theta \leq \frac{\alpha}{2} \\ 0, & -\pi \leq \theta < -\frac{\alpha}{2}, \frac{\alpha}{2} < \theta \leq \pi \end{cases} \quad (15)$$

where for convenience of analysis, here θ ranges from $-\pi$ to π . Note that I_Q only represents the mathematical average of the current waveform if it were not truncated, and thus can be positive and negative. From Fig. 8 and the definition of conduction angle, $I_{max} = I_Q + I_m$, and $I_Q + I_m \cos(\alpha/2) = 0$, thus (15) can be expressed in terms of I_{max} :

$$i_{DS} = \begin{cases} \frac{I_{max}}{1 - \cos \frac{\alpha}{2}} (\cos \theta - \cos \frac{\alpha}{2}), & -\frac{\alpha}{2} \leq \theta \leq \frac{\alpha}{2} \\ 0, & -\pi \leq \theta < -\frac{\alpha}{2}, \frac{\alpha}{2} < \theta \leq \pi \end{cases} \quad (16)$$

The DC and harmonic components of the drain current are obtained by calculating the corresponding coefficients of the Fourier series:

$$I_{DC} = I_0 = \frac{1}{2\pi} \int_{-\frac{\alpha}{2}}^{+\frac{\alpha}{2}} \frac{I_{max}}{1 - \cos \frac{\alpha}{2}} (\cos \theta - \cos \frac{\alpha}{2}) d\theta = \frac{I_{max}}{2\pi} \cdot \frac{2 \sin \frac{\alpha}{2} - \alpha \cos \frac{\alpha}{2}}{1 - \cos \frac{\alpha}{2}} \quad (17)$$

$$\begin{aligned} I_n &= \frac{2}{2\pi} \int_{-\frac{\alpha}{2}}^{+\frac{\alpha}{2}} \frac{I_{max}}{1 - \cos \frac{\alpha}{2}} (\cos \theta - \cos \frac{\alpha}{2}) \cos \theta d\theta \\ &= \frac{I_{max}}{2\pi} \cdot \frac{2}{1 - \cos \frac{\alpha}{2}} \left[\frac{\sin \frac{n-1}{2}\alpha}{n(n-1)} - \frac{\sin \frac{n+1}{2}\alpha}{n(n+1)} \right] \end{aligned} \quad (18)$$

where $n \geq 1$. Maximum power efficiency as a function of conduction angle can be plotted using the results from (17) and the $n = 1$ case (fundamental) from (18). Along with normalized maximum output RF power, this is plotted in Fig. 9. Note that although deep class C operation would yield a maximum efficiency of close to 100%, its output power is also low, which means for a certain overall power gain, class C PAs would need more powerful driver amplifiers, thus the overall efficiency advantage over PAs with larger conduction angles is less than illustrated in the plot. Moreover, class C PAs are nonlinear even with harmonic traps at the output, making them less popular choices in modern communication applications.

II.E. Class D Power Amplifier

As stated before, the active device in a PA can also be used as a switch, resulting in switching mode PAs. Strictly speaking, such circuits should be termed power converters instead of amplifiers, since there

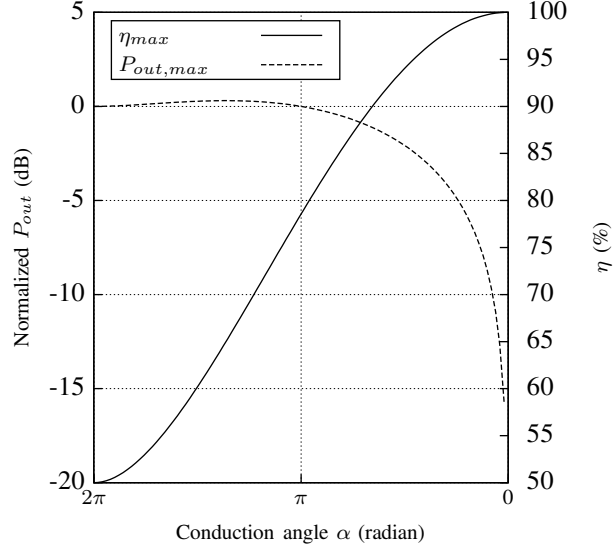


Fig. 9. Maximum efficiency and output power as a function of conduction angle.

is not a strong correlation between the input and output power, they simply convert the DC power from the power supply to RF power.

A straight forward implementation of such an idea is the class D PAs, whose simplified schematic is shown in Fig. 10 [20]. The output series RLC resonator is alternately connected to V_{DD} and ground for each half of the RF cycle, and if the resonator is tuned at the carrier frequency, the current through each switch would be a half-wave sinusoid that complements each other, resulting in a total output current that is a full sinusoid. The voltage and current waveforms are shown in Fig. 11.

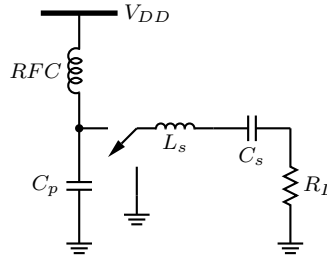


Fig. 10. Conceptual schematic of a class D PA.

Simple calculations of the waveforms reveal that the DC and fundamental components of the square-wave drain voltage are $V_{DD}/2$ and $V_{DD} \cdot 2/\pi$, respectively. Similarly, the DC and fundamental components of the drain current are I_{max}/π and $I_{max}/2$, respectively. The ideal power efficiency of class D is therefore

$$\eta_D = \frac{\frac{1}{2} (I_{max}/2) (V_{DD} \cdot 2/\pi)}{(I_{max}/\pi) (V_{DD}/2)} = 100\% \quad (19)$$

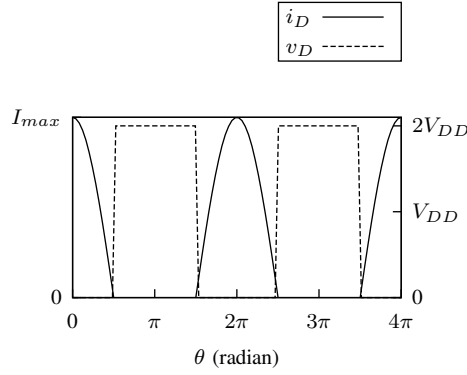


Fig. 11. Class D current and voltage waveforms.

The difficulty of implementing class D PAs for RF applications is the need for complementary switches. The floating switch between V_{DD} and the RLC resonator is usually implemented as a p-type device. Due to the low mobility of holes as compared with electrons, the p-type device is usually two to three times larger in size than the n-type device with a similar power capacity. Not only the loss of p-type device itself would reduce the power efficiency, the large input capacitance requires large driver stage power consumption to accommodate sharp hard switching at RF. The use of transformers between driver and the output switches, with one in-phase and the other anti-phase, makes it possible to implement the two switches using both n-type device [21], but the need of large transformers as well as two power switches still limits its applications to frequencies less than 100 MHz, and therefore an in-depth analysis of class D amplifier is left out here.

II.F. Class E Power Amplifier

The schematic of a class E PA is shown in Fig. 12. The parasitic capacitance at the drain of the transistor can be absorbed into C_p , and package inductance can be absorbed in L_s . Since the transistor is used as a switch, its drain voltage v_D and current i_D waveforms satisfy 1) v_D is negligible when i_D is nonzero and 2) i_D is negligible when v_D is nonzero. The uniqueness of class E PAs is the design of the output filter such that in addition to the two conditions above, ideal operation of class E would result in v_D and i_D waveforms satisfying the following conditions [22]: 3) the rise of v_D at transistor turn-off should be delayed until after the transistor is off, 4) v_D should be brought back to zero at the time of the transistor turn-on, and 5) the slope of v_D should be zero at the time of turn-on. Because of such properties, especially condition the last one regarding to the zero voltage slope, class E PAs are able to achieve high efficiency even the power transistor has finite transition time. The ideal voltage and current

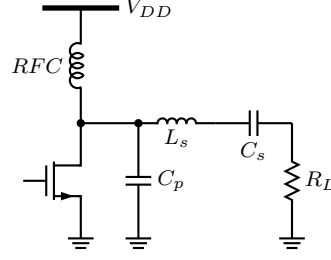


Fig. 12. Conceptual schematic of a class E PA.

waveforms are shown in Fig. 13.

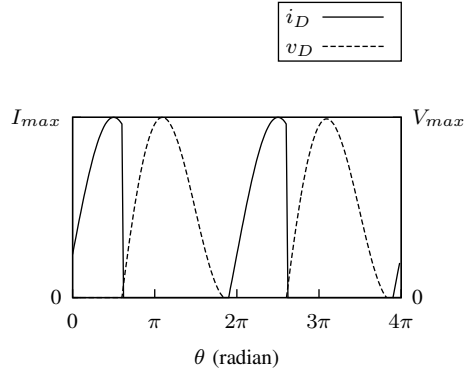


Fig. 13. Class E current and voltage waveforms.

As the case of class D, class E achieves a maximum theoretical efficiency of 100%, but it has several advantages. First, the unavoidable drain capacitance of the transistor causes power loss in class D operations [21], whereas in class E, according to condition 4) above, $v_D = 0$ when the switch is closed, hence there is no switching loss due to charging and discharging of the drain capacitance. This property of class E, also referred as “soft switching” or “zero-voltage switching”, makes it possible for a low-cost switching mode transistor to be used in high power RF applications. Also, because hard switching and square waveforms are not required, class E shows better tolerance of process variations.

The drawback of class E PAs is that the transistor is under large stress. Detailed analysis shows that the maximum drain voltage can be approximately $3V_{DD}$ [22, 23]. This has imposed a serious design constrain for CMOS PAs as the technology scales.

II.G. Class F Power Amplifier

Recall in the analysis of reduced conduction angle PAs, it was assumed that the output filter would attenuate all but the fundamental harmonics. On the other hand, class F PAs intentionally add harmonics at

the output to boost the power efficiency.

Starting with an ideal class B PA, its drain current i_D is a half-wave sinusoid, and its tuned drain voltage v_D is sinusoidal, shown in Fig. 14(a). Its maximum efficiency is 78.5% as calculated before. This efficiency can be improved if v_D can be “flattened”, i.e. it has a sharper transition than sinusoidal and stays at lower value for longer time in the half RF cycle when i_D is nonzero, such as the one shown in Fig. 14(b). This is done by adding odd harmonic components to v_D , which is realized by having infinite impedance at odd harmonics in the output filter. If infinite number of odd harmonics are added to v_D , it becomes a square wave, and the waveforms becomes that of class D, with an efficiency of 100%, Fig. 14(c).

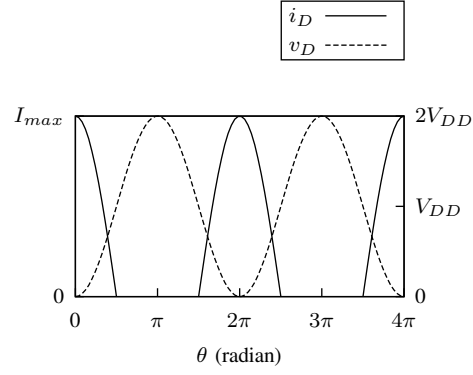
Even without adding all odd harmonics, there would be significant efficiency improvement. For example, consider adding only third harmonic, then the drain volatage is

$$v_D = V_{DD} + V_1 \cos \theta + V_3 \cos 3\theta \quad (20)$$

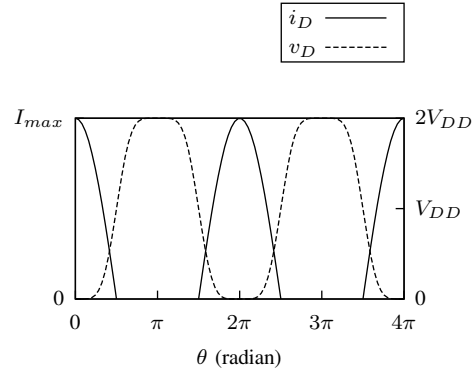
where $\theta = \omega t$. Maximum flatness requires that the second derivative of v_D at $\theta = \pi$ is zero [24]. Combine this with the restraint that v_D cannot exceed $2V_{DD}$, then V_1 and V_3 in (20) are solved to be $V_1 = \frac{9}{8}V_{DD}$, $V_3 = -\frac{1}{8}V_{DD}$. Since the current waveform is the same as that of class B, the efficiency in this case is

$$\eta_F = \frac{1}{2} \frac{V_1 I_{max}/2}{V_{DD} I_{max}/\pi} = \frac{9\pi}{32} \approx 88.4\% \quad (21)$$

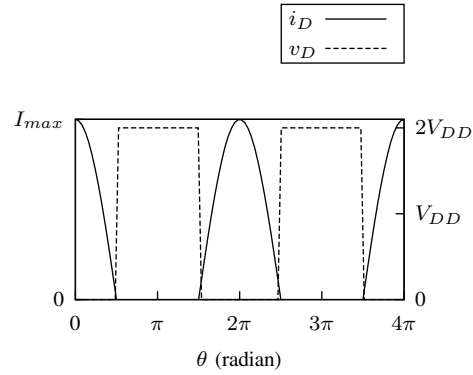
One of the drawbacks of class F is the same as that of class B, which the difficulty in designing an accurate and robust bias scheme. Another disadvantage is that the output filter is more complex, which leads to more power loss. Since the output filter is usually the last stage before the antenna, its power loss is more detrimental to the overall efficiency.



(a)



(b)



(c)

Fig. 14. Class F current and voltage waveform illustrations. (a) Ideal class B current and voltage, (b) class F, realized by adding third and fifth harmonics to the class B drain voltage, and (c) class F with all odd harmonics added to the drain voltage becomes identical to class D.

III. DIGITAL COMMUNICATION SYSTEMS

III.A. Introduction

Modern communication systems widely use digital modulations because of advances in digital signal processing (DSP). In a generic digital RF transmitter which is the focus of this work, information such as voice and image is first digitized then compressed, serialized, and pulse-shaped in the digital domain, before getting converted back to analog form and up-converted to the specific frequency band according to the respective communication standard. The power amplifier (PA) then comes into play and send out the information through the antenna with a certain amount of RF power. The block diagram of such a transmitter is shown in Fig. 15.

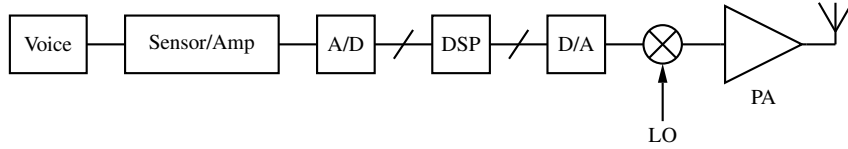


Fig. 15. Block diagram of a digital RF transmitter.

Although this work focuses on the design and implementation of the PA, it is important to have a working understanding of the basic concepts of digital modulations as well as some common communication standards, and this section serves as an overview of these topics.

III.B. Digital Modulation

Digital modulation transfers a digital bit stream over an analog bandpass channel [18]. Due to its digital nature, the modulating signal usually takes one out of two possible values, although in some modulation schemes there can be more than two states. Therefore digital modulation techniques are often termed *keying*, derived from the Morse key used for telegraph.

Different modulation schemes are used in various communication standards, so a brief introduction to them is necessary to reveal some properties of the standards.

III.B.1. Basic Concepts: Starting from the frequency domain, the modulation process is equivalent to moving the complex baseband signal S_{BB} to the carrier frequency f_c :

$$S_{RF}(f) = \frac{1}{2}S_{BB}(f - f_c) + \frac{1}{2}S_{BB}^*(-f - f_c) \quad (22)$$

where $(*)$ denotes complex conjugate. In the time domain, (22) is transformed to be

$$s_{RF}(t) = \frac{1}{2}s_{BB}(t)e^{j\omega_c t} + \frac{1}{2}s_{BB}^*(t)e^{-j\omega_c t} = \Re \{s_{BB}(t)e^{j\omega_c t}\} \quad (23)$$

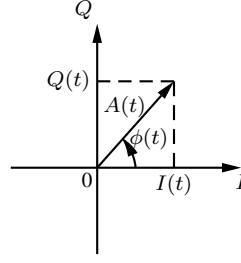


Fig. 16. Relationship between two representations of complex envelope.

where $\omega_c = 2\pi f_c$. The complex signal $s_{BB}(t)$ is called the *complex envelope* of the real signal $s_{RF}(t)$, it can be decomposed into either polar or cartesian form. In polar form, the complex envelope and the real narrowband signals are

$$s_{BB}(t) = A(t)e^{j\phi(t)} \quad (24a)$$

$$s_{RF}(t) = \Re \left\{ A(t)e^{j\phi(t)} e^{j\omega_c t} \right\} = A(t) \cos [\omega_c t + \phi(t)] \quad (24b)$$

where $A(t)$ and $\phi(t)$ are amplitude and phase modulations, respectively. The same process applied to cartesian decomposition would result in

$$s_{BB}(t) = I(t) + jQ(t) \quad (25a)$$

$$s_{RF}(t) = \Re \left\{ [I(t) + jQ(t)] e^{j\omega_c t} \right\} = I(t) \cos \omega_c t - Q(t) \sin \omega_c t \quad (25b)$$

where $I(t)$ and $Q(t)$ are called the *in-phase* and *quadrature* components, respectively. Fig. 16 shows the relationship between the two sets of baseband representations. If the horizontal axis represents the I-component while the vertical represents the Q-component, then

$$A(t) = \sqrt{I^2(t) + Q^2(t)} \quad (26a)$$

$$\phi(t) = \tan^{-1} \frac{Q(t)}{I(t)} \quad (26b)$$

In digital communications, the complex envelope would result in discrete points on the IQ plane, and such plots are termed *constellation diagram*.

The two formulations of the baseband signal are the basis of polar modulation and quadrature modulation, where either $A(t)$ and $\phi(t)$ or $I(t)$ and $Q(t)$ are modulated, respectively. Polar modulation has separate paths for amplitude and phase modulations. The envelope has constant amplitude in the phase modulation path, therefore a switching mode PA can be used. Compared with linear PAs, switching mode

PAs have higher power efficiency, and are less sensitive to antenna impedance variations that are common especially in hand-held electronic devices [25]. However, polar modulation also has its shortcomings. For example, due to AM-PM conversion, the amplitude modulation would inevitably cause phase distortion. In the phase modulation path, although the amplitude of the envelope is constant, since the phase variations are abrupt in digital communication systems, the resultant waveform would show sharp transitions, which in the frequency domain would translate into high-frequency spurs. The implementation of simultaneous amplitude and phase modulation with robust delay mismatch control is difficult and complicated [26], and simulations reveal that polar modulation would lead to worse linearity performance in terms of spectral leakage and error vector magnitude (EVM) when compared with quadrature modulation [27]. On the other hand, simultaneous modulation of amplitude and phase in quadrature modulation systems is not a big issue because, as will be shown, when decomposed into the I and Q components, the phase modulation also shows up as amplitude variations and since quadrature signals are orthogonal, they do not interfere with each other. Also, only one local oscillator would be needed, with an additional 90° phase shifter, and due to the ease with combining from and splitting into two independent component parts of the signal as a result of the symmetrical structure, quadrature modulation is more widely adopted in modern digital communication systems [26].

III.B.2. Phase-shift Keying (PSK) Modulation: A general PSK modulated signal can be expressed as

$$s_{RF}(t) = A \cos [\omega_c t + \phi(t)] = A \cos \phi(t) \cos \omega_c t - A \sin \phi(t) \sin \omega_c t \quad (27)$$

Thus the I and Q components are

$$I(t) = A \cos \phi(t) \quad (28a)$$

$$Q(t) = A \sin \phi(t) \quad (28b)$$

where the amplitude A is constant. As shown in (28), quadrature modulation convert the phase modulation to amplitude modulations in the I and Q components. An M -bit digital system would lead to 2^M possible ϕ values, and to reduce detection error, they are $(2\pi/2^M)$ radians apart.

The simplest case of PSK is binary phase-shift keying (BPSK), where a one-bit system generates two phases, $\phi = 0$ and $\phi = \pi$. From (28), for BPSK $(I, Q) = (\pm A, 0)$, and thus the contellation diagram should ideally be that of Fig. 17.

Quadrature phase-shift keying (QPSK) is a more common modulation scheme, because at the same

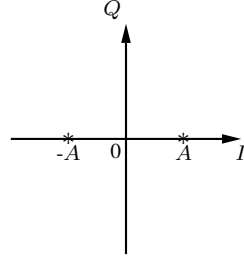


Fig. 17. Ideal BPSK constellation diagram.

symbol rate, it is able to transmit twice as much information as that of BPSK. Adding more number of bits would further increase the bandwidth efficiency, but as symbols getting close on the constellation diagram, the system becomes more prone to errors.

A common choice of ϕ for QPSK and the corresponding (I, Q) pairs are summarized in Table II, and the constellation diagram is shown in Fig. 18. Note that with such a choice of phases, the QPSK can be viewed as a sum of an I-channel BPSK and a Q-channel BPSK.

TABLE II
QPSK Format Summary

Dibit	ϕ	I	Q
00	$\pi/4$	$A/\sqrt{2}$	$A/\sqrt{2}$
01	$-\pi/4$	$A/\sqrt{2}$	$-A/\sqrt{2}$
10	$3\pi/4$	$-A/\sqrt{2}$	$A/\sqrt{2}$
11	$-3\pi/4$	$-A/\sqrt{2}$	$-A/\sqrt{2}$

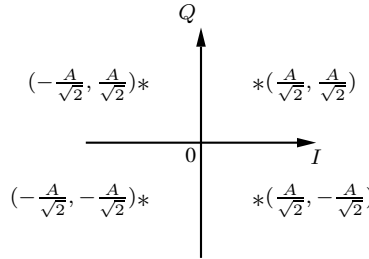


Fig. 18. Ideal QPSK constellation diagram.

III.B.3. Quadrature Amplitude Modulation (QAM): As its name suggests, QAM is a modulation scheme where amplitude modulation is applied on both in-phase and quadrature channels, i.e.

$$s_{RF}(t) = I_m \cos \omega_c t - Q_m \sin \omega_c t \quad (29)$$

where I_m and Q_m usually both take on M values and thus there are altogether M^2 possible constellation points. For example, a 4-bit scheme can be transmitted using 16QAM, because two bits would result in

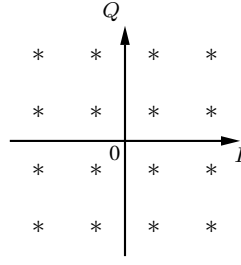


Fig. 19. Ideal 16QAM constellation diagram.

four levels in I_m , and the other two would generate four levels in Q_m . The constellation diagram of a 16QAM is shown in Fig. 19.

III.B.4. Orthogonal Frequency Division Multiplexing (OFDM): The modulation schemes described so far all modulate the digital bit series into a single carrier. OFDM is a method of modulating the digital information into multiple carriers. In this scheme, the channel bandwidth is divided into multiple subchannels, each of which independently modulated. The symbol rate of each subchannel is set to be the reciprocal of the subchannel frequency spacing Δf , so all subchannels are orthogonal to each other [18].

The major advantage of OFDM is that it is less sensitive to transmission media characteristics. This is because as the subchannel bandwidth becomes sufficiently small, the medium frequency response can be approximated to be flat, and according to Nyquist theorem, the system is free of intersymbol interference (ISI). Consequently, OFDM is a common choice of modulation format in communication systems where the attenuation of the channel is severe or there is multipath propagation. The disadvantage of OFDM modulation is the resultant high PAPR signals would require linear transceiver, which would usually lead to inferior power efficiency.

III.C. Wireless Communication Standards

Some of the most common communication standards are introduced here. Among the numerous specifications laid out in each standard, only those that are most closely related to PA design are discussed here.

III.C.1. Global System for Mobile Communications (GSM): As a replacement for first-generation (1G) analog cellular networks, GSM is the first cellular phone standard that is based on digital modulations [2]. The GSM uses Gaussian minimum-shift keying (GMSK) modulation scheme, which is a variation of PSK with Gaussian pulse shape. To increase user capacity, it applies time-division multiple access (TDMA) signalling and frequency-division duplexing (FDD). During each time slot up to eight users can

TABLE III
GSM PA Specifications

System	Uplink (MHz)	$P_{out,max}$ (dBm)
GSM 850	824.2 - 849.2	33
GSM 900	880.0 - 915.0	33
GSM 1800	1710 - 1785	30
GSM 1900	1850 - 1910	30

be accommodated and the data rate per user is 270 kb/s. In an FDD system, the transmission and reception of the signals are at different frequencies, thus there is good isolation between the receiver and transmitter. Since in mobile devices the transmitter and receiver are usually built at close vicinity, even on the same chip, FDD is employed in many wireless communication systems [28]. Extension of GSM to facilitate data communications led to the developments of general packet radio services (GPRS) and enhanced data rates for GSM evolution (EDGE), both considered second-and-half generation (2.5G) standards.

The PA design for GSM applications does not require amplitude linearity because it uses constant-envelope modulations. Therefore switching mode PAs can be used, and the main design effort is targeted at power efficiency optimization. GSM standard is integrated by the third-generation partnership project (3GPP) for backward compatibility and the full set of specifications can be found in [29], Table III summarizes the key specifications of GSM that are related to PA design.

III.C.2. Wireless Local Area Network (WLAN): Based on the IEEE 802.11 standards, WLAN uses OFDM modulation to reduce sensitivity to multipath effects and has a high data rate. Unlike GSM, WLAN applies time-division duplexing (TDD). The advantage of TDD is that the transmission and reception of the signal use the same frequency, thus direct communications between two transceivers, which is an important feature in WLAN, is made much easier to realize. The major drawback of TDD is that unintended but strong transmission signals may appear at the same frequency at the receiver, and thus the linearity requirements for both the transmitter and receiver are stringent.

Due to the use of OFDM modulation, linear PA is required for WLAN transmitter design. Therefore both efficiency and linearity are important design targets to tackle. Furthermore, since OFDM modulation usually has a high PAPR (about 10 dB), the power efficiency at the back-off region determines the average efficiency. The complete specifications of WLAN can be found in [30], and the PA-related specifications of IEEE 802.11g is listed in Table IV.

III.C.3. Wideband Code Division Multiple Access (WCDMA): WCDMA is the primary cellular standard for the third-generation (3G) wireless communications and the most commonly used member of the

TABLE IV
IEEE 802.11g PA Specifications

Specification	Value
Frequency (MHz)	2412 - 2484
No. of carriers	52
Channel bandwidth (MHz)	22
Max. P_{out} (dBm)	20
Max. EVM (dB)	-25
	-20 @ 10 MHz offset
	-28 @ 20 MHz offset
ACLR (dBc)	-40 @ 30 MHz offset

universal mobile telecommunications system (UMTS). It uses FDD scheme, and accommodates up to ten carriers.

Similar to the case of WLAN, WCDMA standard has high requirements in regard to linearity [31]. Also because of multicarrier modulation, it is vital to ensure the power back-off efficiency is high. The specifications related to PA design are listed in Table V.

TABLE V
WCDMA PA Specifications

Specification	Value
Uplink (MHz)	1900 - 1980
No. of carriers	10
Channel bandwidth (MHz)	3.84
Max. P_{out} (dBm)	30
Max. EVM (dB)	-15
	-33 @ 5 MHz offset
ACLR (dBc)	-43 @ 10 MHz offset

IV. NONLINEAR EFFECTS OF POWER AMPLIFIER

IV.A. Introduction

Conventional modulations, such as frequency modulation (FM), frequency-shift keying (FSK), and Gaussian minimum-shift keying (GMSK) have their information stored in the frequency variations, not the envelope amplitude, and thus do not require linear amplification. If the envelope amplitude does contain information and thus varies, then linear amplification is required [19]. Of course amplitude modulations fall into this category, but phase modulation (PM), phase-shift keying (PSK), and multicarrier modulations such as orthogonal frequency-division multiplexing (OFDM) also have envelope amplitude variations that need to be preserved throughout the transceiver chain, therefore linear power amplifiers (PAs) are required.

As the available frequency bands for wireless communications have become increasingly crowded, bandwidth efficiency has become an important design considerations. As a result, most modern communication systems use raised-cosine (RC) pulse shaping. RC filter's spectrum can be shaped arbitrarily close to rectangulars, whose bandwidth efficiency is maximum, while maintaining minimal intersymbol interference (ISI) [18]. Therefore the required transmitter pulse shape is root-raised-cosine (RRC), which results in a envelope with a peak to average power ratio (PAPR) of 3-6 dB, depending on the specific modulation applied [19].

OFDM has found its popularity in wideband digital communications such as wireless local access network (WLAN), digital television (DTV), and fourth generation long term evolution (4G LTE), due to its multi-carrier nature that enables more reliable transmission and reception compared with single-carrier schemes [18]. The resultant envelope's PAPR falls in the range of 8-13 dB [19].

In either of the modulation schemes aforementioned that have non-constant envelopes, PA's linearity is a crucial design specification. Lack of linearity would result in distortion of the amplified signal, which results in the frequency domain as unwanted components in frequencies other than the designated frequency bands. This phenomenon is also called spectrum regrowth, and the worst case of it usually happens in the adjacent channel. Therefore the adjacent channel leakage ratio (ACLR), defined as the ratio of the adjacent power to the main channel power, is usually one of the toughest test of PA's linearity.

This section will discuss the MOSFET nonlinearity mechanisms, as well as how they are manifested in single-tone, two-tone, and multi-tone tests. Also, AM-PM conversion will be introduced.

IV.B. MOSFET Nonlinearity Mechanisms

Semiconductor transistors always exhibit nonlinearities to some extent. Fig. 20 illustrates a typical I-V characteristic of a MOSFET [20]. The transistor conducts negligible current in the cutoff region as

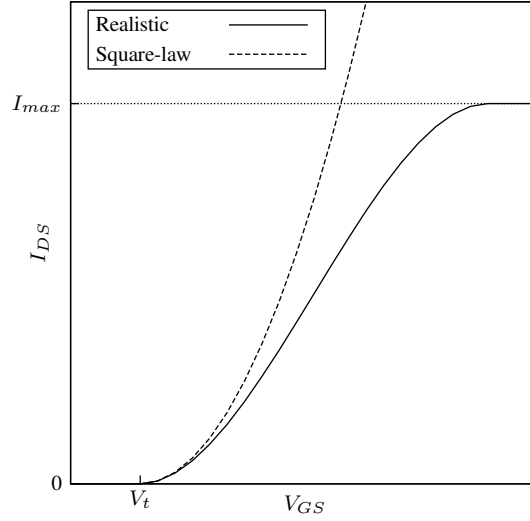


Fig. 20. Illustrative I-V characteristic of a MOSFET.

$V_{GS} < V_t$, due to the absence of a conducting channel. As the channel is formed after V_t , the transistor enters active region, the $I_{DS} - V_{GS}$ relation would follow a quadratic equation if long channel model of the MOSFET is used, i.e.

$$I_{DS} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_t)^2 \quad (30)$$

for an NMOS transistor, where μ_n is the average electron mobility in the channel, C_{ox} is the gate oxide capacitance per unit area, W and L are the width and length of the channel, respectively, and V_t is the threshold voltage.

It can be seen in Fig. 20 that it does not take long for the I-V characteristics to deviate from the square law. This is the result of several second-order effects [32]:

a) *Channel-length modulation*: In the active region, $V_{DS} > V_{GS} - V_t$, or $V_{GD} < V_t$. In other words, the gate-drain voltage is not large enough to sustain the conduction channel, or the channel is “pinched off”. As a result, there exists a depletion region between the pinch-off end of the channel and the drain, which would cut into the effective channel length. This effect can be modeled by adding a V_{DS} -dependent

term to (30):

$$I_{DS} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_t)^2 (1 + \lambda V_{DS}) \quad (31)$$

where λ is proportional to the effective channel length, to the first-degree approximation. According to (31), this effect is approximately proportional to the drain current and inversely proportional to the channel length. Therefore the deviation after V_t is apparent when deep sub-micron CMOS transistors in the RF PA applications are delivering a large amount of current. As V_{GS} increases, V_{DS} would decrease, and hence the depletion region would shrink, resulting in a larger effective channel length, which in turn causes the drain current to be less than that predicted by the square law.

b) Velocity saturation: At relatively low V_{GS} after V_t , the V_{DS} of the transistor is large. Combined with a short channel length, the electrical field between drain and source becomes very strong, and the electron drift velocity becomes less proportional to the electrical field as in the low-field case and tend to saturate. A first-order approximation reveals

$$v_d = \frac{\mu_n \mathcal{E}}{1 + \mathcal{E}/\mathcal{E}_c} \quad (32)$$

where \mathcal{E} is the electrical field, and \mathcal{E}_c is termed the critical field. As a result, the drain current in the active region becomes

$$I_{DS} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} V_{DS(act)}^2 = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left\{ \mathcal{E}_c L \left[\sqrt{1 + \frac{2(V_{GS} - V_t)}{\mathcal{E}_c L}} - 1 \right] \right\}^2 \quad (33)$$

where the $V_{DS(act)}$ term can be approximated by

$$V_{DS(act)} = \mathcal{E}_c L \left[\sqrt{1 + \frac{2(V_{GS} - V_t)}{\mathcal{E}_c L}} - 1 \right] \approx (V_{GS} - V_t) \left(1 - \frac{V_{GS} - V_t}{2\mathcal{E}_c L} \right) \quad (34)$$

In the presence of a strong drain-to-source electrical field, $\mathcal{E}_c L$ is relatively small, and according to (34) and (33), I_{DS} would also be less than that predicted by the square law.

c) Mobility degradation: As V_{GS} further increases, the vertical electrical field originated from the gate voltage causes the carriers in the channel to be closer to the silicon surface, where surface imperfections would impede their movement. This effect can be modeled as a degradation of the carrier mobility [33], where the effective mobility is given by

$$\mu_{eff} = \frac{\mu_n}{1 + \theta (V_{GS} - V_t)} \quad (35)$$

where the parameter θ is inversely proportional to the oxide thickness. As MOS technology scales down, the oxide thickness shrinks, and thus this effect also causes I_{DS} reductions.

In addition to the effects mentioned above, as V_{GS} further increases, V_{DS} would continue to decrease, and the transistor would enter triode region. The drain current would then be determined by the load of the transistor, and flatten out to the saturated value I_{max} as the transistor enters the deep triode region.

Refer back to Fig. 20, if the transistor enters cutoff or deep triode region, there will be strongly nonlinear effects due to clipping. On the other hand, the PA is driven to its maximal current capacity, indicating a good efficiency performance. The switch mode PAs operate in strong nonlinear regions, resulting in higher efficiency. However, as mentioned before, modern communication schemes usually have very stringent linearity standards, requiring most of the PAs operating between the two extremes in Fig. 20, sometimes referred to as the weakly nonlinear region, thus the analyses below would focus in this region.

In the weakly nonlinear region, the PA's input and output can be related by a power series

$$i_{out} = \sum_{i=1}^{\infty} G_i v_{in}^i \quad (36)$$

where G_i are complex coefficients, representing both amplitude and phase nonlinearities, and they include the memory effects that devices under high frequency excitations usually exhibit. The Volterra series are formulated in this way and serve as a rigorous nonlinearity analysis tool for RF applications [34, 35]. However, such a formulation becomes too complicated as the transistor models become more sophisticated, and hence, gives less design insights compared to simpler series. Also, for modern communication systems that require high linearity, a power series of up to third-order generally gives most of the required information, assuming the PA operates up to the 1-dB compression point [36–38]. Analysis in this paper utilizes a real-valued power series up to the third order component. Under this simplification, (36) becomes

$$i_{out} = g_1 v_{in} + g_2 v_{in}^2 + g_3 v_{in}^3 \quad (37)$$

where g_1 represents the linear transconductance gain of the PA, and g_2 and g_3 represent the second- and third-order distortion terms, respectively.

IV.C. Single-tone Test: Harmonic Distortion and Gain Compression

IV.C.1. Harmonic Distortion: Consider a voltage signal with constant amplitude v_a and frequency f_0 , also known as a continuous wave (CW), is input to a nonlinear PA. In other words, $v_{in} = v_a \cos(2\pi f_0 t)$

in (37). Using the identities $\cos^2 \theta = \frac{1}{2} (1 + \cos 2\theta)$ and $\cos^3 \theta = \frac{1}{4} (3 \cos \theta + \cos 3\theta)$, (37) becomes

$$i_{out} = g_1 v_a \cos \theta_0 + \frac{1}{2} g_2 v_a^2 (1 + \cos 2\theta_0) + \frac{1}{4} g_3 v_a^3 (3 \cos \theta_0 + \cos 3\theta_0) \quad (38)$$

For the sake of conciseness, $\theta_0 = 2\pi f_0 t = \omega_0 t$. One obvious consequence of a CW signal passing through a nonlinear PA is that components at multiples of the carrier frequency are generated, and they are termed harmonic tones. Generally speaking, even-order distortions would generate even-order harmonics up to their orders, while odd-order distortions would generate odd-order harmonics up to their orders. For instance, as resulted in (38), second-order distortion generates a DC term and a second harmonic, while third-order distortion generates a distortion term at the fundamental tone, in addition to a third harmonic.

Second harmonic distortion (HD_2) is defined as the ratio of the amplitude of the output signal component at the second harmonic to that of the fundamental. Therefore according to the nonlinear model in (37),

$$HD_2 = \frac{1}{2} \left| \frac{g_2}{g_1} \right| v_a \quad (39)$$

Third harmonic distortion (HD_3) can be defined similarly, and in the current model,

$$HD_3 = \frac{1}{4} \left| \frac{g_3}{g_1} \right| v_a^2 \quad (40)$$

Note in (40) that it is assumed the distortion at the fundamental contributed by the g_3 is negligible, which is valid in the weakly nonlinear region. In the weakly nonlinear region, HD_2 and HD_3 are usually very small and are more conveniently expressed in decibel:

$$HD_i|_{dB} = 20 \log HD_i \quad (41)$$

where $i = 2, 3, \dots$. For RF applications, harmonic distortions do not give very realistic indications of the system linearity. This is because the harmonic tones are at multiples of the fundamental tone, which are very far apart. In most of the narrow-band RF systems, the output filter would reduce the harmonic tones, so the harmonic distortions may appear very small even if the system is substantially nonlinear.

IV.C.2. Gain Compression: Collecting terms at the fundamental in (38) yields

$$i_{out, fund} = \left(g_1 + \frac{3}{4} g_3 v_a^2 \right) v_a \cos \theta_0 = \left(g_1 + \frac{3}{4} g_3 v_a^2 \right) v_{in} \quad (42)$$

which indicates that the gain at the fundamental tone may be greater or less than the gain without distortion. These phenomena are termed gain expansion and gain compression, respectively. For MOSFET, usually

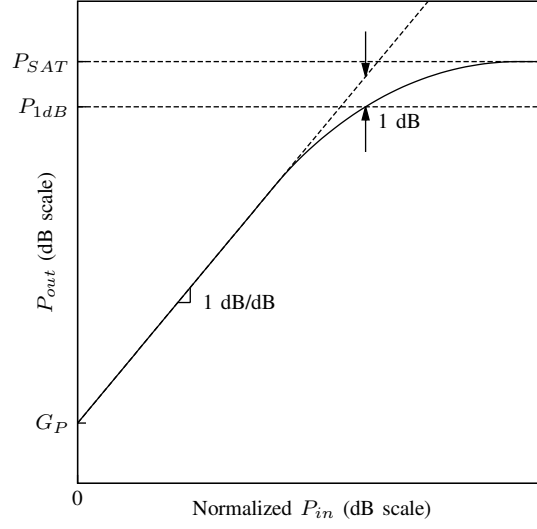


Fig. 21. Typical power transfer characteristics of a PA.

g_1 and g_3 have opposite signs, and would exhibit gain compression. For BJT, whose I-V relationship is exponential, g_1 and g_3 have the same sign, which would show gain expansion [20]. However, the gain expansion of BJT would only happen at a close vicinity around the operating point, because as the input voltage amplitude increases, higher order distortion contributions also come into play, and more importantly, as shown in Fig. 20, the output current for any practical system would saturate at high input levels, preventing any possible gain expansions.

Therefore the typical power transfer characteristic of a practical PA would look like the one shown in Fig. 21. At low P_{in} levels, the PA is mostly linear, i.e. $P_{out} = G_P P_{in}$, where G_P denotes the linear power gain. On the dB-dB scale plot, the linear portion of the transfer curve would appear to be a straight line with a slope of 1 dB/dB and y -intercept of G_P . As P_{in} increases, the power gain starts to compress due to nonlinear, especially odd-order, distortions, and hence P_{out} starts to deviate from the ideal linear characteristics, shown in the dashed line. In RF applications, the gain compression is quantified by the *1-dB compression point*, defined as the point at which the power gain has dropped by 1 dB from its small-signal asymptotic value, it can be referred to the input or the output power. For transmitter building blocks, the output-referred 1-dB compression point is often used, whereas for receiver building blocks, input-referred 1-dB compression point is more common. As P_{in} further increases, most practical PA would saturate, reaching the maximum RF power it can generate, P_{SAT} . P_{SAT} indicates the PA's power capability, but for linear PAs, P_{1dB} is usually viewed as the upper limit of the PA's dynamic range.

According to the definition of P_{1dB} and (42), for a third-order distortion PA, the input amplitude that corresponds to P_{1dB} satisfies

$$20 \log \left| g_1 + \frac{3}{4} g_3 v_{ic}^2 \right| = 20 \log |g_1| - 1 \quad (43)$$

which leads to

$$v_{ic} = \sqrt{0.145 \left| \frac{g_1}{g_3} \right|} \quad (44)$$

In summary, under the single-tone condition, second-order distortion results in a signal-dependent DC offset and a harmonic distortion term that appears at twice the fundamental frequency, both of which can be easily filtered out by the output filter of a PA. On the other hand, the third-order distortion would lead to a non-linear signal dependent term at the fundamental frequency in addition to the third-order harmonic distortion. The distortion term at the fundamental frequency is considered the cause of gain compression [28]; hence, in terms of linearity considerations, third-order distortion is more critical when designing RF PA.

IV.D. Two-tone Test: Intermodulations and Intercept Point

IV.D.1. Intermodulations: Now consider an input signal that consists of two tones, that is,

$$v_{in} = v_1 \cos \omega_1 t + v_2 \cos \omega_2 t \quad (45)$$

The nonlinear PA described so far would generate harmonic tones at $2\omega_i$ and $3\omega_i$, $i = 1, 2$, and in addition, due to the presence of two tones, there will be nonlinear tones at linear combinations of their frequencies. This nonlinearity is termed intermodulation (IM).

a) Second-order Intermodulations: The second-order IM products are

$$IM_2 = g_2 v_1 v_2 [\cos(\theta_1 + \theta_2) + \cos(\theta_1 - \theta_2)] \quad (46)$$

where again to reduce unnecessary complexity, $\theta_i = \omega_i t$, $i = 1, 2$. If the two input tones are close to each other, then the two terms in (46) are at approximately twice the carrier frequency and at close to DC. In a typical RF PA application, IM_2 are usually filtered out by the output filter, and do not contribute to severe nonlinearity problems to the system. Also due to the filtering effects, IM_2 does not serve as a practical linearity metric for RF PAs.

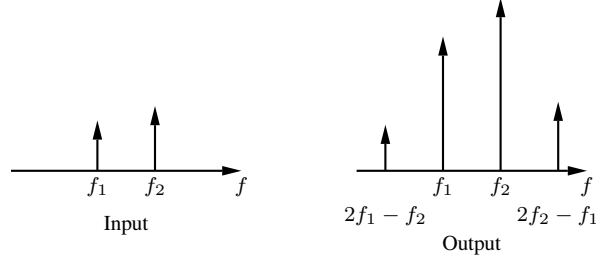


Fig. 22. Third-order intermodulation.

b) Third-order Intermodulations: The third-order IM products are

$$IM_3 = \frac{3}{4}g_3 [v_1^2 v_2 \cos(2\theta_1 \pm \theta_2) + v_2^2 v_1 \cos(2\theta_2 \pm \theta_1)] \quad (47)$$

The $(2\omega_1 + \omega_2)$ and $(2\omega_2 + \omega_1)$ terms would appear close to three times the carrier frequency, if the two input tones are again assumed to be close. Therefore these terms are again of less importance. However, $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$ are very close to the input tones: they are just outside of ω_1 and ω_2 by the frequency difference of the two main tones, as illustrated in Fig. 22. Defined by the ratio of the IM_3 product at $(2\omega_1 - \omega_2)$ to the fundamental at f_1 or IM_3 at $(2\omega_2 - \omega_1)$ to fundamental at f_2 , the third-order intermodulation distortion is therefore calculated according to (47) and (38):

$$IMD_3 = \frac{3}{4} \left| \frac{g_3}{g_1} \right| v_1 v_2 \quad (48)$$

IV.D.2. Intercept Point: Consider the two input tones to have the same amplitude v_a , then according to (47) the IM_3 amplitude becomes $\frac{3}{4}g_3 v_a^3$, whereas by comparison, the fundamental tones would have an amplitude of $g_1 v_a$. Although IM_3 products are less than fundamental tone amplitude when v_a is small, they increase as a function of v_a^3 , while the fundamentals increase in proportion to v_a . The third-order intercept point (IP_3) is defined to be where IM_3 and fundamental have the same amplitude. IP_2 can be similarly defined, but since IM_2 is rarely used in PA applications, so is IP_2 .

Note that IP_3 is a term that can indicate a system's linearity, and thus provides some design insights, usually it is not a quantity that can be directly measured, as the case of P_{1dB} . This is because as v_a increases, the gain compresses and higher order distortions become significant. IP_3 is usually obtained by extrapolation of measured data. Theoretical IP_3 calculation would simply set the IM_3 product and fundamental have the same amplitude, assuming the two input tones have the same amplitude v_a , then from (38) and (47),

$$v_{iip3} = \sqrt{\frac{4}{3} \left| \frac{g_1}{g_3} \right|} \quad (49)$$

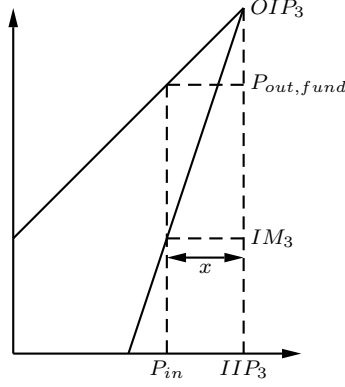


Fig. 23. Geometrical interpretation of IP_3 calculation.

Compare IP_3 with other linearity metrics that describe the third-order distortion such as IM_3 or IMD_3 in (47) and (48), it is clear that the advantage of IP_3 is that it is independent of input signal level. This makes it a popular quantity to compare linearity performance of different RF circuits. Compare v_{iip3} with that of P_{1dB} in (43),

$$IP_3 - P_{1dB} = 10 \log \left(\frac{4}{3} \left| \frac{g_1}{g_3} \right| \right) - 10 \log \left(0.145 \left| \frac{g_1}{g_3} \right| \right) \approx 9.64 \text{ dB} \quad (50)$$

Therefore as a rule of thumb, IP_3 is higher than P_{1dB} by about 10 dB. To get IP_3 from simulations, sweeping two large-signal tones and extrapolating would consume a large amount of computing resources. One way to reduce simulation time is to perform only one two-tone simulation, in which the amplitudes of the two input tones are identical, and they are set to be small enough that both the fundamental and third-order distortion terms are in their asymptotic ranges, but large enough so they can be measured accurately. Then the input-referred IP_3 can be calculated as

$$IIP_3 = P_{in} + \frac{1}{2} |IMD_3| \quad (51)$$

where all quantities are in dB-scale, and P_{in} is the input power of one of the two input tones. The derivation of (51) can be geometrically explained in Fig. 23. On logarithmic scales, the fundamental and third-order power transfer characteristic lines would have slopes of 1 dB/dB and 3 dB/dB, respectively. Therefore from Fig. 23, $OIP_3 - P_{out,fund} = x$, and $OIP_3 - IM_3 = 3x$, with x being the power difference between IIP_3 and P_{in} . Recall the definition of $IMD_3 = IM_3 - P_{out,fund}$, it can be easily deduced that $x = |IMD_3|/2$, hence comes the result in (51).

Another way to simulate IP_3 without consuming a large amount of simulation resources is to use only one large tone in the two-tone test and sweep this tone only, while the other tone is kept as a small-signal

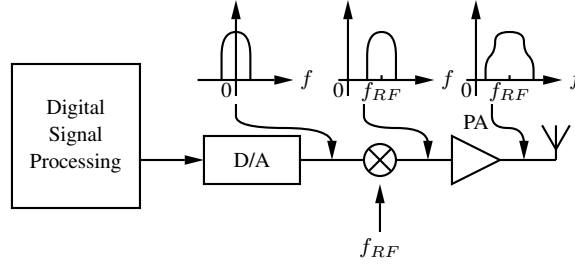


Fig. 24. Typical RF transmitter.

tone [39]. The caveat of this approach is that in IP_3 calculations, IM_3 and the corresponding fundamental tone should not be adjacent, i.e. $(2f_1 - f_2)$ and f_2 , $(2f_2 - f_1)$ and f_1 tone amplitudes are to be set equal in extrapolating the transfer curves to get IP_3 , which is apparent from (47), because doing so would cancel out the non-equal tone effects and come to the same result as in (49).

IV.E. Multitone Nonlinear Effects

Linearity is one of the major concerns in current and near future wireless communication systems. Of all the requirements related to linearity in wireless communication standards, the requirements on spectral leakage, often in the form of adjacent-channel leakage ratio (ACLR), are usually the most demanding specifications for radio frequency integrated circuits (RFIC) design.

A typical RF transmitter block diagram is conceptually shown in Fig. 24. Modern communication standards usually require a certain form of filtering applied to the digital information in order to limit the bandwidth. However, due to nonlinearity of the RF front-end, the shape of the input signal is not preserved, which means the spectrum is not limited to the desired bandwidth. This effect is termed *spectral regrowth* [28] and is quantified by ACLR, which is defined as the ratio of the integrated power in the adjacent channel to the power in the transmitted channel [31].

To assess ACLR performance in the design phase, envelope simulations are needed [39]. However, such a simulation is very expensive in terms of computing resources and simulation time, which is especially true for RF PA designs because of the large number of active devices involved. Moreover, due to the random nature of signals used in communications, it is not easy to analytically relate ACLR to PA design parameters. Multitone signals are similar to the band-limited signals used in communication systems in the frequency domain, but its simulation is still relatively expensive, and depending on the number of tones at the input source, it may not be well-supported by the simulator [40].

This subsection presents an analysis that relates a multitone ACLR to two-tone third-order intermodulation distortion (IMD_3) of a nonlinear system, which enables a quick estimate of the PA spectral regrowth

using a fast two-tone simulation in the early design phase.

Consider an input signal that consists of N identical tones evenly spaced in the frequency domain:

$$v_{in} = v_a \sum_{n=1}^N \cos [\omega_0 + (n-1) \Delta\omega] t \quad (52)$$

where $\Delta\omega$ denotes the angular frequency spacing. Rearranging (52) reveals more clearly that the normalized input resembles a modulated signal:

$$\begin{aligned} \frac{v_{in}}{v_a} &= \frac{1}{2} \left(\sum_{n=1}^N \exp \{j [\omega_0 + (n-1) \Delta\omega] t\} \right. \\ &\quad \left. + \sum_{n=1}^N \exp \{j [-\omega_0 - (n-1) \Delta\omega] t\} \right) \\ &= \frac{1}{2} \left(e^{j\omega_0 t} \frac{1 - e^{jN\Delta\omega t}}{1 - e^{j\Delta\omega t}} + e^{-j\omega_0 t} \frac{1 - e^{-jN\Delta\omega t}}{1 - e^{-j\Delta\omega t}} \right) \\ &= \frac{1}{2} \left(e^{j\omega_0 t} \frac{e^{jN\frac{\Delta\omega}{2}t}}{e^{j\frac{\Delta\omega}{2}t}} \frac{e^{-jN\frac{\Delta\omega}{2}t}}{e^{-j\frac{\Delta\omega}{2}t}} - \frac{e^{jN\frac{\Delta\omega}{2}t}}{e^{j\frac{\Delta\omega}{2}t}} \right. \\ &\quad \left. + e^{-j\omega_0 t} \frac{e^{-jN\frac{\Delta\omega}{2}t}}{e^{-j\frac{\Delta\omega}{2}t}} \frac{e^{jN\frac{\Delta\omega}{2}t}}{e^{j\frac{\Delta\omega}{2}t}} - \frac{e^{-jN\frac{\Delta\omega}{2}t}}{e^{-j\frac{\Delta\omega}{2}t}} \right) \\ &= \frac{\sin(N\frac{\Delta\omega}{2}t)}{\sin(\frac{\Delta\omega}{2}t)} \cos \left(\omega_0 + \frac{N-1}{2} \Delta\omega \right) t \end{aligned} \quad (53)$$

Clearly $(\omega_0 + \frac{N-1}{2} \Delta\omega)$ is the center frequency of the frequency band from ω_0 to $(\omega_0 + (N-1)\Delta\omega)$, which can be viewed as the carrier frequency, and hence the N -tone signal would have an “envelope” of

$$v_{env} = v_a \frac{\sin(N\frac{\Delta\omega}{2}t)}{\sin(\frac{\Delta\omega}{2}t)} \quad (54)$$

The envelope voltage has maxima at $t = n\frac{2\pi}{\Delta\omega}$, $n = 0, 1, 2, \dots$. For instance, using the L'Hôpital's rule, it can be shown that

$$v_{env,max} = \lim_{t \rightarrow 0} v_a \frac{\sin(N\frac{\Delta\omega}{2}t)}{\sin(\frac{\Delta\omega}{2}t)} = N v_a \quad (55)$$

To calculate the root-mean-square (RMS) value of v_{in} composed of N equal-magnitude tones, the Parseval's identity is used, leading to

$$v_{in,rms} = \sqrt{2v_a^2 \sum_{n=1}^N \left(\frac{1}{2}\right)^2} = \sqrt{\frac{N}{2}} v_a \quad (56)$$

Combining the results from (55) and (56), the PAPR of an N -tone signal is computed as

$$PAPR = \frac{P_{max}}{P_{avg}} = \frac{v_{env,max}^2/2}{v_{in,rms}^2} = N \quad (57)$$

where a normalized resistance of 1Ω is assumed. In other words, (57) reveals that, for example, a modulated signal with a PAPR of 10 dB can be emulated by a 10-tone signal. Although the frequency spectrum of a multitone signal may have a similar profile to a modulated signal with similar PAPR, it is still quite different from the continuous band-limited signal's spectrum due to the discrete nature of its spectrum. This is evident when the signals go through a nonlinear system, and spectral regrowth is observed. As will be shown later, even when two signals show similar ACLR, the spectra can be different in shape.

For wireless communications, the exact definition of adjacent channel, i.e., the frequency offset and integration bandwidth, depend on the specific communication format, but for multitone analysis, one can assume that the adjacent channel starts just outside of the main channel, and has the same bandwidth as that of the main channel. The second-order distortion would not contribute to tones in the adjacent channels for high-IF narrow-band signals; therefore, only third-order distortions are considered in analyzing multitone adjacent channel powers (ACP). Substituting v_{in} from (52), the third-order distortion output is

$$i_{out3} = g_3 v_{in}^3 = g_3 v_a^3 \left(\sum_{k=1}^N \cos \omega_k t \right)^3 \quad (58)$$

Next, the amplitude of each tone in the adjacent channel is calculated, and the ACP is the sum of their output power. To simplify the analysis, it is assumed the PA operates in the weakly nonlinear regime, and thus the distortion terms that fall within the passband are assumed to be much less than that of the amplified signal, and therefore are not accounted. Also, due to symmetry, only the upper ACP is calculated, namely the tones at frequencies ω_N up to ω_{2N-1} . An extensive analysis was performed in [37], but phase uncorrelation was assumed, i.e., it was assumed that each distortion tone outside of the passband is contributed to by nonlinear terms with uncorrelated phases, and thus the distortion amplitude was computed by vectorially adding all the nonlinear contributions. Although the frequency response of the PA might result in different phases, the difference is small in most current communication standards. Therefore, analytical results using the approach in [37] should be considered the best-case scenario. In this work, the analysis considers the worst-case condition, in which all distortion terms are considered in-phase. Therefore, for each tone in the adjacent channel, the output currents that were contributed from

all possible distortion mechanisms are added before the result gets squared to calculate the power.

Observation of (58) leads to dividing i_{out3} components into three categories: where $\omega_k = \omega_l = \omega_m$, $\omega_k = \omega_l \neq \omega_m$, and $\omega_k \neq \omega_l \neq \omega_m$.

a) *Harmonic terms* ($k = l = m$): The output current tones due to third-order distortion are

$$i_{HD3} = M_1 g_3 v_a^3 \sum_{k=1}^N \cos^3 \omega_k t \quad (59)$$

where M_1 denotes the multinomial coefficient, which is unity in this case since all three summations in (58) need to contribute the term with the same ω_k . It can be shown that (59) would result in small tones whose effects should affect the magnitude and phase of the inband tones and then degrading the quality of the final constellation; those effects, however, are not the focus of this paper. These components also result in third-order harmonic components, which by narrowband assumption are out of the adjacent channel range. Therefore these harmonic distortion terms are not considered in this analysis.

b) *Intermodulation terms* ($k = l \neq m$): The output current tones as a result of intermodulation are

$$i_{IM3} = M_2 g_3 v_a^3 \sum_{l=1}^N \sum_{\substack{m=1 \\ m \neq l}}^N \cos^2 \omega_l t \cos \omega_m t \quad (60)$$

The multinomial coefficient in this case is $M_2 = C(3, 2) = 3$ because for each pair of (l, m) , the coefficient M_2 is the combination of choosing 2 (that contribute ω_l) out of the three summation terms in (58). Each term in (60) would expand as below:

$$\begin{aligned} \cos^2 \omega_l t \cos \omega_m t = \frac{1}{2} \left[\cos \omega_m t + \frac{1}{2} \cos (2\omega_l + \omega_m) t \right. \\ \left. + \frac{1}{2} \cos (2\omega_l - \omega_m) t \right] \end{aligned} \quad (61)$$

The first term in (61) results in tones that are in-band, and the tones as the result of the second term are out of band. The only term that contributes to the upper adjacent channel spurs is the third term that satisfies $\omega_l > \omega_m$, $\frac{N}{2} < l \leq N$, and thus, the tone positions are $\omega_{IM3} = 2\omega_l - \omega_m = \omega_l + n_{lm}\Delta\omega$, where n_{lm} is the index difference between the two tones ω_l and ω_m . Therefore, for ω_{IM3} to appear in the adjacent channel, the tone position difference between ω_l and ω_m needs to be greater than that between ω_l and the upper end of the channel ω_N , or $n_{lm} > n_{Nl}$. If $l = N$, all $1 \leq m \leq N - 1$ would result in an ω_{IM3} that is in the adjacent channel, at $\omega_{N+1}, \omega_{N+2}, \dots, \omega_{2N-1}$. If $l = N - 1$, $m = N - 2$ would not result in an ω_{IM3} that is in the adjacent channel, so $m_{max} = N - 3$, and the resulting adjacent channel tones are at $\omega_{N+1}, \omega_{N+2}, \dots, \omega_{2N-3}$. This process of analysis is summarized in Table VI.

TABLE VI
Intermodulation Adjacent Tone Position Analysis

l	m_{max}	$\omega_{IM3,max}$ index	No. of tones in adj. ch.
N	$N - 1$	$2N - 1$	$N - 1$
$N - 1$	$N - 3$	$2N - 3$	$N - 3$
\vdots	\vdots	\vdots	\vdots
$N - i$	$N - 2i - 1$	$2N - 2i - 1$	$N - 2i - 1$
\vdots	\vdots	\vdots	\vdots
$\frac{N+1}{2} + 1$ (N odd)	2	$N + 2$	2
$\frac{N}{2} + 1$ (N even)	1	$N + 1$	1

Counting occurrences at each tone in the upper adjacent channel from Table VI reveals that at ω_{N+p} , the occurrence of IM_3 distortion term $K_{IM3,N+p}$ is

$$K_{IM3,N+p} = \lceil \frac{N-p}{2} \rceil \quad (62)$$

where $\lceil x \rceil$ denotes the *ceiling function* of x , which is defined as the least integer greater than or equal to x [41], and $p = 1, 2, \dots, N - 1$. Therefore, the distortion due to IM_3 at tones in the upper adjacent channel would result in

$$i_{IM3,adj} = \frac{3}{4} g_3 v_a^3 \sum_{p=1}^{N-1} \lceil \frac{N-p}{2} \rceil \cos \omega_{N+p} t \quad (63)$$

c) *Triple-beat terms* ($k \neq l \neq m$): The third-order distortion output in this category is

$$i_{TB} = M_3 g_3 v_a^3 \sum_{\substack{k,l,m=1 \\ k \neq l \neq m}}^N \cos \omega_k t \cos \omega_l t \cos \omega_m t \quad (64)$$

These distortion terms are called *triple beat* [42], and since each tone in the triple beat is different, the multinomial coefficient $M_3 = P(3, 3) = 3! = 6$. Expansion of each term in (64) yields

$$\begin{aligned} \cos \theta_k \cos \theta_l \cos \theta_m &= \frac{1}{4} [\cos (\theta_k + \theta_l + \theta_m) \\ &\quad + \cos (\theta_k + \theta_l - \theta_m) \\ &\quad + \cos (\theta_k - \theta_l + \theta_m) \\ &\quad + \cos (\theta_k - \theta_l - \theta_m)] \end{aligned} \quad (65)$$

where $\theta_i = \omega_i t$, $i = k, l, m$. Without loss of generality, assume $k > l > m$, then with the narrowband assumption, in (65), only the second term has distortion terms that fall in the upper adjacent channel. The first term is approximately three times the passband frequency; the third term corresponds to frequencies

TABLE VII
Triple-beat Adjacent Tone Position Analysis

k	No. terms at ω_{N+1}	No. terms at ω_{N+2}	$\omega_{TB,max}$ index
N	$N-2$	$N-3$	$2N-2$
$N-1$	$N-4$	$N-5$	$2N-4$
\vdots	\vdots	\vdots	\vdots
$N-i$	$N-2i-2$	$N-2i-3$	$2N-2i-2$
\vdots	\vdots	\vdots	\vdots
$\frac{N}{2} + 2$ (N even)	2	1	$N+2$
$\frac{N-1}{2} + 2$ (N odd)	1	0	$N+1$

at $(\omega_k - n_{lm}\Delta\omega)$, but since $k > l > m$, these frequencies all fall in the passband, and the last term results in negative frequencies, but their absolute values are frequencies at $(\omega_m - n_{kl}\Delta\omega)$, which are either in the passband or in the lower adjacent channel.

Observation of the second term in (65) reveals that the frequencies of this distortion term are at $(\omega_k + n_{lm}\Delta\omega)$. If $k = N$, then $1 \leq m < l \leq N-1$, leading to $(N-1-1 = N-2)$ pairs of (l, m) such that $n_{lm} = 1$, and consequently there would be a distortion term at ω_{N+1} . Similarly, there are $(N-1-2 = N-3)$ pairs of (l, m) that result in a distortion contribution at ω_{N+2} , and so on. Finally, the farthest index “distance” between ω_l and ω_m , in the case of $k = N$, is $(N-2)$, so $\omega_{TB,max} = \omega_{2N-2}$. When $k = N-1$, $1 \leq m < l \leq N-2$, there would be $(N-2-2 = N-4)$ pairs of (l, m) such that $n_{lm} = 2$ and thus a distortion contribution at $\omega_{TB} = \omega_{N+1}$, and $\omega_{TB,max} = \omega_{2N-4}$. Similar analysis can be carried out for other k values, and the results are summarized in Table VII.

The occurrence of a triple-beat at ω_{N+p} is then computed as

$$K_{TB,N+p} = \lfloor \frac{N-p}{2} \rfloor \lceil \frac{N-p}{2} \rceil \quad (66)$$

where $\lfloor x \rfloor$ is the *floor function* of x , which is defined as the greatest integer less than or equal to x [41], and $p = 1, 2, \dots, N-2$. Therefore, the triple beat in the upper adjacent channel would result in an output current of

$$i_{TB,adj} = \frac{3}{2} g_3 v_a^3 \sum_{p=1}^{N-2} \lfloor \frac{N-p}{2} \rfloor \lceil \frac{N-p}{2} \rceil \cos \omega_{N+p} t \quad (67)$$

As mentioned before, if assume the distortion contributions to each adjacent tones are all in-phase, then the nonlinear output in the upper adjacent channel is the sum of results from (63) and (67):

$$i_{adj,u} = \frac{3}{4} g_3 v_a^3 \sum_{p=1}^{N-1} \left[\lceil \frac{N-p}{2} \rceil \left(1 + 2 \lfloor \frac{N-p}{2} \rfloor \right) \right] \cos \omega_{N+p} t \quad (68)$$

It can be shown that after some mathematical manipulations and a change of variable, (68) can be simplified to

$$i_{adj,u} = \frac{3}{4} g_3 v_a^3 \sum_{p=1}^{N-1} \left[\frac{p(p+1)}{2} \right] \cos \omega_{2N-p} t \quad (69)$$

And the adjacent channel power then becomes the sum of power in each tone, assuming that the tones *themselves* are all uncorrelated.

$$\begin{aligned} P_{adj} &= \frac{1}{2} \left(\frac{3}{4} g_3 v_a^3 \right)^2 \sum_{p=1}^{N-1} \frac{1}{4} (p^2 + p)^2 \\ &= \frac{1}{2} \left(\frac{3}{4} g_3 v_a^3 \right)^2 \frac{1}{4} \left[\frac{1}{5} (N-1)^5 + (N-1)^4 \right. \\ &\quad \left. + \frac{5}{3} (N-1)^3 + (N-1)^2 + \frac{2}{15} (N-1) \right] \\ &= \frac{1}{2} \left(\frac{3}{4} g_3 v_a^3 \right)^2 F(N) \end{aligned} \quad (70)$$

where

$$\begin{aligned} F(N) &= \frac{1}{4} \left[\frac{1}{5} (N-1)^5 + (N-1)^4 \right. \\ &\quad \left. + \frac{5}{3} (N-1)^3 + (N-1)^2 + \frac{2}{15} (N-1) \right] \end{aligned}$$

The passband output power is

$$P_{ch} = \frac{N}{2} (g_1 v_a)^2 \quad (71)$$

If a two-tone signal with equal amplitude of v_a is applied to the PA, the IMD_3 , expressed in decibel, is

$$IMD_3|_{dB} = 10 \log \left(\frac{3}{4} \left| \frac{g_3}{g_1} \right| v_a^2 \right)^2 \quad (72)$$

The ACLR of an N -tone signal that is applied to the PA can then be obtained from (70) and (71), and more importantly, it can be related to a two-tone IMD_3 described by (72):

$$\begin{aligned} ACLR &= 10 \log \frac{P_{adj}}{P_{ch}} \\ &= IMD_3 - 10 \log N + 10 \log F(N) \end{aligned} \quad (73)$$

Analysis in [37] assumes that the phases of all distortion components are random. Such an assumption

provides the lower limit of ACLR:

$$ACLR_{\min} = IMD_3 - 10 \log N + 10 \log G(N) \quad (74)$$

where

$$G(N) = \frac{1}{12} [4N(N^2 - 1) - 3(N^2 - N \bmod 2)]$$

Therefore, in the design phase, a simple two-tone simulation can be used to get IMD_3 , and from (73) and (74), one can obtain a basic idea of the multitone ACLR range, which indicates what could happen in the corresponding ACLR as a result from a modulated input signal that has a similar PAPR.

To verify the analysis of this work, an RF power amplifier was designed and fabricated in TSMC 40 nm CMOS, and Fig. 25 shows the microphotograph of the chip. The linear PA operates at 1.9 GHz, with a measured continuous-wave (CW) saturated output power $P_{SAT} \approx 35$ dBm and power gain of 38 dB. The PA itself is designed with three stages, and the last two stages can be switched to improve power efficiency, but for the purpose of verifying the concept in this paper, the switching functionality is not activated in the following simulations and measurements. For more details of the design and implementation of the power amplifier, the readers are referred to [43].

First, a set of simulations are carried out to verify the relationship between multitone ACLR and the number of tones. The input source to the PA provides a multitone signal, and the Fourier analysis is performed at the PA output to calculate the ACLR. The input signal's total bandwidth is kept constant, and the amplitude of the individual input tones, v_a , is kept the same as the number of tones N is varied from 2 to 9. The value of v_a is chosen such that in case when the maximum number of tones is reached, in this case 9, the peak input voltage would not exceed the PA's saturation limit. Based on the two-tone test and thus the corresponding IMD_3 , (73) can be computed and compared to simulation results. The results are shown in Fig. 26. Note that as a reference, the ACLR predicted by (74) is also included in the plot. As expected the simulation results are better than the results of (73) but worse than that predicted by (74). As previously mentioned, the purpose of this analysis is to provide a quick estimate of the ACLR at the early design phase; therefore, a prediction of the ACLR range should suffice. For large N s, the difference between the results predicted by (73) and (74) can be approximated by taking the dominant terms in $F(N)$ and $G(N)$ only. Even when $N = 10$, for instance, the ACLR should be within a 10 dB window, which is manageable in the early design phase, especially when considering that the time and resources it takes to get the estimate is fairly little.

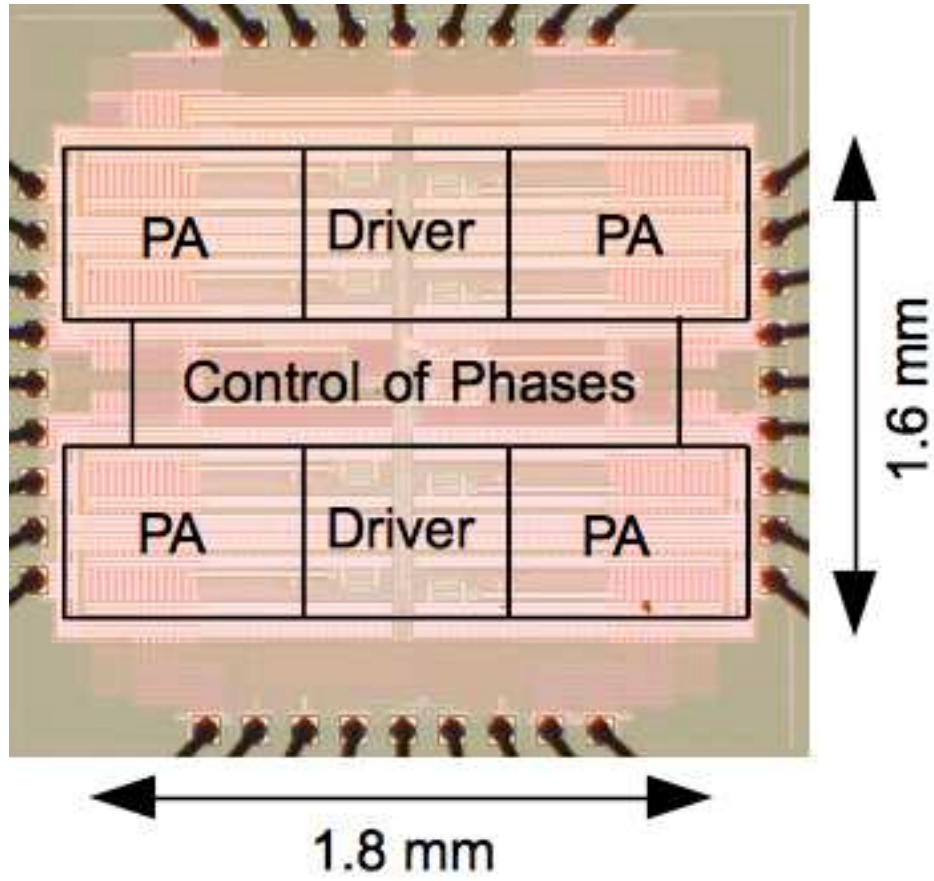


Fig. 25. Microphotograph of the chip.

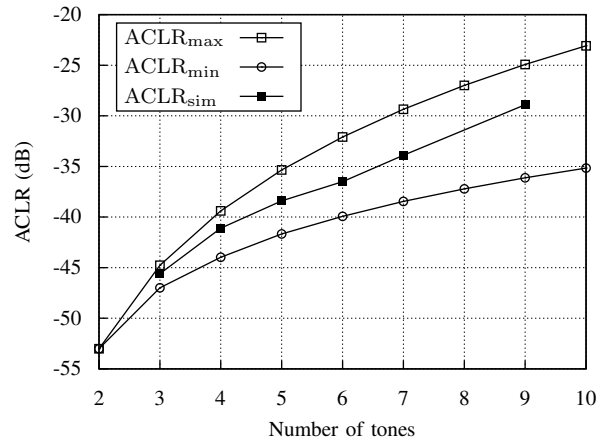


Fig. 26. Simulated and predicted multitone ACLR as a function of number of tones.

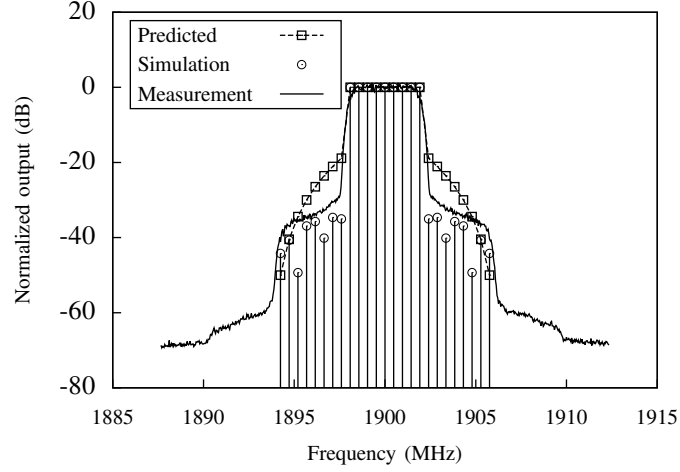


Fig. 27. Multitone output spectrum.

As a second testbed, an OFDM signal with a PAPR of about 9 dB is applied to the fabricated PA and the output power spectrum is measured. According to (57), a 9-tone signal has a similar PAPR; therefore, the normalized output spectrum calculated according to (69) and the corresponding simulation results are compared with the measured PA output spectrum, as shown in Fig. 27.

As expected, the theoretical calculations from (69) and (73) represent the worst case, and due to the phase response of the PA and its output matching network, the simulation results give a better nonlinearity performance. Although similar to PAPR, the multitone and its corresponding OFDM spectrum are different in shape, mainly due to the difference between their time domain envelope. Nevertheless, the ACLR turns out to be similar, which again validates the proposed methodology. The simulated 9-tone ACLR is about -30 dBc, whereas the measurement result is -33 dBc ACLR for the OFDM signal.

The results from simulations and measurements show the validity of the analysis. Therefore, as a design guideline from a linearity perspective, a two-tone simulation can provide some indications of the PA linearity in the presence of multitone or even modulated signals. For a specific target modulation format with a known PAPR, the designers can then offer an educated estimation of the ACLR from a quick two-tone simulation. If the PA has a strong memory effect, the IM_3 would become a function of the frequency spacing of the two tones, and are asymmetric [20]. The cause of memory effect and the means to reduce it are out of the scope of discussion in this paper, but to account for the memory effect, a design margin of about 5 dB in ACLR should be allocated.

IV.F. AM-to-PM Conversion

Previously described nonlinearities can be summarized as amplitude modulation to amplitude modulation (AM-AM) conversion, i.e. amplitude modulation at the input of the PA (or any other nonlinear system) results in disproportional output amplitude modulation. In frequency domain, this phenomenon can be interpreted as the envelope of the modulated signal having additional frequency components at the output, as manifested by previously described intermodulation and triple beat.

Another phenomenon of nonlinearity is amplitude modulation to phase modulation (AM-PM) conversion, in which the PA's phase response becomes a function of the input amplitude. Nonlinear capacitance of the PA transistor and high- Q matching networks are all possible causes of AM-PM conversion [20, 44, 45]. However, AM-PM conversion is only dominant beyond the 1-dB compression point [20].

To analyze AM-PM conversion, consider an input signal $v_{in} = \cos \theta_m \cos \theta$, where for simplicity, $\theta = \omega t$, and the input amplitude is normalized to unity. Here ω_m represent the envelope, and ω is the carrier, and the signal can be decomposed into two tones, one at $(\omega - \omega_m)$ and the other at $(\omega + \omega_m)$. As stated, AM-PM conversion is a function of the input amplitude, and therefore in this case, there would be an additional phase Ψ added to the output of the PA, and Ψ has a period that is half of the envelope. In the simplest model, assume

$$\Psi = \frac{\phi}{2} (1 + \cos 2\theta_m) \quad (75)$$

that is the normalized phase error has a maximum of ϕ , average of $\phi/2$, and a frequency of $2\omega_m$. Assume the phase error is small, i.e. $\cos \Psi \approx 1$, $\sin \Psi \approx \Psi$, $\cos \frac{\phi}{2} \approx 1$, $\sin \frac{\phi}{2} \approx \frac{\phi}{2}$, and for simplicity, ignore the amplification of the amplitude, then the output is

$$\begin{aligned} v_{out} &= \cos \theta_m \cos (\theta + \Psi) \\ &\approx \cos \theta_m (\cos \theta - \Psi \sin \theta) \\ &= \cos \theta_m \left(\cos \theta - \frac{\phi}{2} - \frac{\phi}{2} \sin \theta \cos 2\theta_m \right) \\ &\approx \cos \theta_m \left\{ \cos \left(\theta + \frac{\phi}{2} \right) - \frac{\phi}{4} [\sin (\theta + 2\theta_m) + \sin (\theta - 2\theta_m)] \right\} \\ &= \cos \theta_m \cos \left(\theta + \frac{\phi}{2} \right) - \frac{\phi}{8} [\sin (\theta \pm \theta_m) + \sin (\theta \pm 3\theta_m)] \end{aligned} \quad (76)$$

Therefore because of the amplitude-dependence of phase distortion, the AM-PM conversion would generate tones at the two input tones at $(\omega \pm \omega_m)$ as well as at the third-order intermodulation tones at $(\omega \pm 3\omega_m)$, and they would vectorically add to the distortions caused by AM-AM conversion. Although derived using a

rather simple model, the above AM-PM results would hold for more complex Ψ and modulation schemes.

V. IMPEDANCE MATCHING NETWORK DESIGN

V.A. Introduction

At RF or microwave frequencies, the electromagnetic wavelength becomes comparable to the PCB traces or even traces on-chip. Wave properties, such as reflections, can be significant if there is impedance mismatch [46]. A direct connection between the PA's output and the antenna is also impractical. If the input impedance of the antenna is $50\ \Omega$, a simple calculation shows that the voltage swing would have to be $10\ V_{pk}$ that corresponds to an output power of 30 dBm, or 1 W. Advance CMOS technology is not able to handle such a voltage swing, hence an impedance matching network that transforms the load impedance at the antenna to a lower value at the output port of the PA is needed. Note that input impedance matching is also needed for measurement purposes, but if the PA is part of a transceiver SoC that has other parts that precede it, then input matching may not be necessary.

At microwave frequencies, transmission lines can be used to implement impedance matching, whereas at low RF, lumped LC networks are more common due to the otherwise large size of transmission lines. At intermediate or high RF, a hybrid approach can be used.

Basic concepts of impedance matching are reviewed in this section, followed by the review of several basic impedance matching network. The discussion would focus on PA output impedance matching, although many of the principles apply to input matching as well.

V.B. Impedance Matching Theory

The PA output impedance matching can be conceptually illustrated in Fig. 28. Here R_L represents the load impedance, which could be the input impedance of the antenna or measurement probe. Usually it is real and has a value of $50\ \Omega$ for RF applications or $75\ \Omega$ in TV systems. Through the impedance matching network, the PA's output is instead terminated at R_T , the termination impedance. Since PA is one of the most expensive devices in the transceiver in terms of power consumption and area, R_T is usually designed for optimal (maximal) output power, therefore its reactance is zero or negligible [20]. Load-pull technique is utilized to determine the R_T value for PA [47], while S-parameter techniques are used for determining R_T to optimize other parameters, such as noise, gain, and stability, for small-signal amplifiers [46, 48].

There are several considerations to be taken into account when designing a matching network. First, the relationship between R_L and R_T needs to be considered. In general, R_T can be greater than or less than R_L , which could result in different matching topologies. Even within the PA design applications, where R_T is usually less than R_L , the ratio between the two would determine how many stages of matching

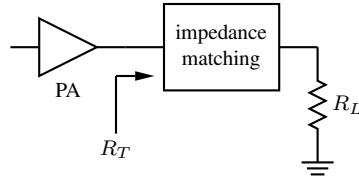


Fig. 28. PA output impedance matching block diagram.

should be used. This is because, as will be shown, that if a single-stage matching is used, then there is no degree of freedom in terms of the quality factor Q of the matching network. As a result, especially when R_L is much larger than R_T , a single-stage matching network would have a very high Q , which in turn would make the matching quite narrow-band. A matching network that has a narrow bandwidth may be sensitive to process and component variations. As can be shown later, the matching network has wider bandwidth with more stages.

In addition to matching the impedance at fundamental frequency, harmonic impedances also need to be well terminated. Usually because of the lowpass nature of the impedance matching network, the harmonics are filtered out. But in case of bands that have stringent harmonic requirements, especially bands that are involved in carrier aggregation [49, 50], then certain harmonic traps, either in the form of a series harmonic open or shunt harmonic short, may be required to be embedded into the matching network.

The other consideration is the insertion loss of the impedance matching circuit. Generally speaking, the less number of component used, the less the loss is. Therefore, the insertion loss trades off with the bandwidth of the matching circuit. Also due to this consideration, the impedance matching network mainly consists of low-loss components, such as inductors, capacitors, and transmission lines.

At microwave frequencies, transmission lines can be used to implement impedance matching, whereas at low RF, lumped LC networks are more common due to the otherwise large size of transmission lines. At intermediate or high RF, a hybrid approach can be used.

V.C. L-match Network

One of the most basic forms of the impedance matching network consists of a shunt element and a series element, forming an L-shaped impedance transformer. For PA design applications, where R_L is usually greater than R_T , there is usually an element lumped at the load impedance R_L , and another in series. Two L-match networks implemented with inductor and capacitor are illustrated in Fig. 29.

In principle, to realize a certain impedance transformation, the component values for both versions of the L-match networks shown in Fig. 29 are the same, and they should have identical in-band performance.

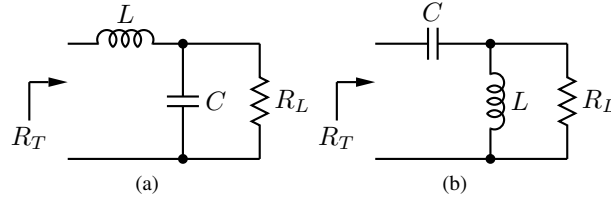


Fig. 29. Lumped-element L-match network, in (a) low-pass and (b) high-pass forms.

But usually the low-pass version is preferred. This is because of the often need for rejection or attenuation at harmonic frequencies, which are all higher than the fundamental tone. Another advantage of the low-pass L-match, especially if a hybrid approach is to be implemented on PCB, is that the inductors can be conveniently replaced by microstrip lines, formed simply by PCB traces over the ground plane, and thus having the low-pass version would enable one to easily change the position of the capacitors along the lines to “tweak” the network, making it a more flexible approach in practice.

To analyze the lumped version of the L-match impedance transformation network, refer back to the low-pass version in Fig. 29(a). The parallel combination of R_L and C results in an impedance of

$$Z_{RC} = \frac{R_L}{1 + j\omega R_L C} = \frac{R_L}{1 + \left(\frac{R_L}{X_C}\right)^2} - jX_C \frac{\left(\frac{R_L}{X_C}\right)^2}{1 + \left(\frac{R_L}{X_C}\right)^2} \quad (77)$$

Therefore the inductance and capacitance are determined such that the real part of Z_{RC} is the desirable termination resistance, R_T , and the imaginary part of Z_{RC} is resonated out by the inductor. Define the impedance transformation factor $m = R_L/R_T$, then according to (77),

$$X_L = R_T \sqrt{m - 1} \quad (78a)$$

$$X_C = \frac{R_L}{\sqrt{m - 1}} \quad (78b)$$

The quality factor Q of the impedance matching network is also an important quantity to take into considerations during design. For a resonant circuit, its Q is equal to the ratio of the resonant frequency to the 3-dB bandwidth:

$$Q = \frac{f_0}{BW} \quad (79)$$

where f_0 denotes the resonance frequency. In other words, Q is inversely proportional to the network bandwidth. Unlike resonators, the impedance matching network is driven by a source: either a real signal source for the input-match case, or the active device that can be modeled as a source for the output-match case. Therefore, in impedance matching applications, a concept of “loaded Q ” is introduced when

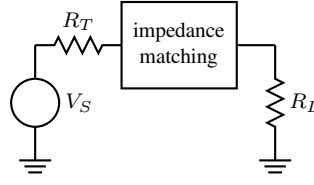


Fig. 30. Concept of loaded Q .

discussing the bandwidth of the network. The loaded Q , Q_L , is defined as the Q near the matching frequency of the impedance matching network driven by a source with the proper (intended) source impedance, as shown in Fig. 30.

For the lumped L-C L-match, the Q of the L-C network can be calculated as

$$Q = \frac{R_L}{X_C} = \sqrt{m - 1} \quad (80)$$

So if the circuit is driven by a voltage source with a series resistance of R_T as shown in Fig. 31, the equivalent resistance of the network is doubled at resonance, and thus

$$Q_L = \frac{1}{2}Q = \frac{1}{2}\sqrt{m - 1} \quad (81)$$

Note that if a lumped L-C matching topology is chosen, then once the impedance ratio m is determined, so is the Q . Since R_L is usually set at 50 Ω , and R_T is chosen to achieve best power or efficiency performance of the PA, the lumped L-C matching network is lack of a degree of freedom to design for the Q .

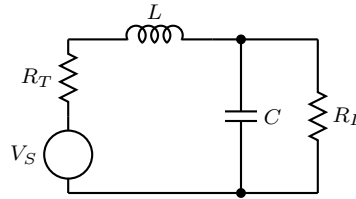


Fig. 31. Lumped L-match driven by the intended source resistance.

The hybrid version of the L-match network has a transmission line in place of the inductor, shown in Fig. 32. G_T and G_L represent the termination and load conductance, respectively. Y_X is the admittance of the parallel R-C network, Y_0 and d are the characteristic admittance and length of the transmission line, respectively.

As hinted in Fig. 32, the analysis would be less involved if performed in terms of admittance, conductance, and susceptance. The impedance transformation strategy of this scheme is to use the lumped

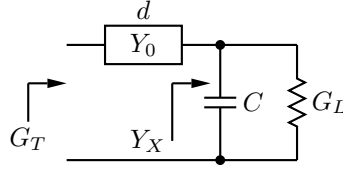


Fig. 32. Hybrid L-match network.

capacitor to change the admittance seen at the end of the transmission line, in order to let the reflection coefficient Γ_X at the right end of the line have the same magnitude as that of G_T . Then the length of the transmission line is determined such that through the length of d , the phase of Γ_X would change to that of Γ_T .

Define $n = G_T/Y_0$, then according to the strategy outlined above,

$$\begin{aligned}
 |\Gamma_X| &= |\Gamma_T|, \text{ or} \\
 \left| \frac{Y_0 - G_L - jB_C}{Y_0 + G_L + jB_C} \right| &= \left| \frac{Y_0 - G_T}{Y_0 + G_T} \right| = \frac{n-1}{n+1}, \text{ or} \\
 B_C &= G_L \sqrt{\frac{Y_0}{G_L} \frac{n^2+1}{n} - \left(1 + \frac{Y_0^2}{G_L^2}\right)}
 \end{aligned} \tag{82}$$

And the length of the transmission line is determined by

$$\begin{aligned}
 \Gamma_X &= \frac{Y_0 - Y_X}{Y_0 + Y_X} = \frac{Y_0 - G_L - jB_C}{Y_0 + G_L + jB_C} \\
 d &= \frac{\lambda}{4\pi} \angle \Gamma_X = \frac{\lambda}{4\pi} \tan^{-1} \left(\frac{2B_C}{G_L \frac{n^2+1}{n} - 2Y_0} \right)
 \end{aligned} \tag{83}$$

where $\lambda = c_{eff}/f$ is the wavelength of the electromagnetic wave, and c_{eff} is the effective propagation speed of light. Similar to (80), the quality factor of this network can be calculated by

$$Q = \frac{B_C}{G_L} = \sqrt{\frac{Y_0}{G_L} \frac{n^2+1}{n} - \left(1 + \frac{Y_0^2}{G_L^2}\right)} \tag{84}$$

From the same logic in the lumped version, the loaded Q is half of the network Q . This expression is a little more complicated than that of the lumped-element case, but for the special case where $G_L = Y_0$, i.e. the transmission line is matched to the load impedance, (84) reduces to

$$Q = \sqrt{n - 2 - \frac{1}{n}} \tag{85}$$

For large n , (80) and (85) converge. In other words, the bandwidth improvement is negligible for power amplifier output impedance matching applications if the characteristic impedance of the transmission line

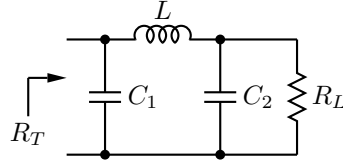


Fig. 33. Π -match network in output matching applications.

is the same as the load termination. Therefore lower characteristic impedance is usually implemented to achieve higher bandwidth. Coincidentally, transmission lines with low characteristic impedance are usually wide, which result in a lower parasitic resistance and thus less power loss and better current handling capacity.

In other words, the hybrid impedance matching network provides the missing degree of freedom to control the quality factor. For example, suppose an impedance matching network is needed to transform $50\ \Omega$ load impedance to $5\ \Omega$. If a lumped-element L-match is used, then according to (80), the Q is 3, and cannot be changed because it is determined only by the impedance transformation ratio. If the hybrid L-match is chosen, and assume the transmission line's characteristic impedance is $50\ \Omega$, then according to (85), the Q is calculated to be 2.8. Furthermore, if the characteristic impedance of the transmission line can be varied, then the Q can be changed accordingly. If the transmission line is implemented such that the characteristic impedance is $30\ \Omega$, then from (84), the Q is 2.55. Compare this result with that of lumped implementation, the network bandwidth can be increased by 15%.

V.D. *Pi-match Network*

The output transistor of power amplifiers usually exhibits a relatively large output parasitic capacitance. In addition, the bond wires connecting the output node on chip to the package have inductance that cannot be ignored at RF. Therefore a simple L-match design would result in inaccurate impedance transformation due to lack of such considerations. An additional capacitor C_1 is added at the end of the inductor, as shown in Fig. 33, forming a Π -match network. The output capacitance of the PA transistor can be absorbed into C_1 , whereas package and PCB trace capacitance can be absorbed into C_2 .

Compared with L-match, the additional element in the Π -match network provides flexibility in choosing the loaded Q , and also results in more matchable range on the Smith chart [46]. Therefore Π -match network is widely used in input-, output-, and interstage-matching applications. As will be shown later, L-match can be treated as a special case of Π -match with one of the capacitance being zero. Therefore the following analysis of the Π -match network start with a more generalized form as shown in Fig. 34,

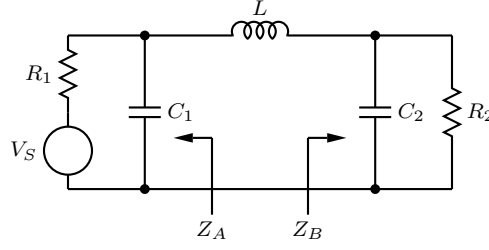


Fig. 34. General schematic for analysis of a Π -match network.

in which the Π -match network is going to match R_1 and R_2 at a certain frequency, and is driven by a voltage source V_S . In the case of input-matching applications, V_S can be the RF source and R_1 the source impedance, and R_2 is the input impedance of the RF circuit; in the case of output-matching applications, V_S can be viewed as the Thevenin equivalent of the PA, with R_1 being the optimal output impedance of the PA and R_2 being the load impedance.

Due to the additional degree of freedom, now the loaded Q of the network can be specified when designing the Π -match, and so the following analysis assumes R_1 , R_2 , Q_L , and the frequency of interest are all specified, and would try to derive design equations to determine L , C_1 , and C_2 .

First introduce

$$B_{C1} = \omega C_1, B_{C2} = \omega C_2, X_L = \omega L \quad (86)$$

and define

$$Q_1 = B_{C1}R_1, Q_2 = B_{C2}R_2 \quad (87)$$

As shown in Fig. 34, define Z_A and Z_B as the parallel equivalent impedance of $\{R_1, C_1\}$ and $\{R_2, C_2\}$, respectively, therefore

$$Z_A = R_A - jX_A, Z_B = R_B - jX_B \quad (88)$$

Basic circuit analysis can show that

$$R_A = \frac{R_1}{1 + Q_1^2}, X_A = R_A Q_1 \quad (89a)$$

$$R_B = \frac{R_2}{1 + Q_2^2}, X_B = R_B Q_2 \quad (89b)$$

At conjugate match, $R_A = R_B$, $X_L = X_A + X_B$, and since the loaded Q at resonance can be expressed as $Q_L = \frac{X_L}{R_A + R_B}$, the following relationships can be derived from (89):

$$Q_L = \frac{1}{2} (Q_1 + Q_2) \quad (90)$$

$$\frac{R_1}{R_2} = \frac{1 + Q_1^2}{1 + Q_2^2} \quad (91)$$

$$X_L = R_1 \frac{2Q_L}{1 + Q_1^2} = R_2 \frac{2Q_L}{1 + Q_2^2} \quad (92)$$

From (90) and (91), the condition on which the Π -match is designable can be derived. For example, rearranging (91) for an expression of Q_2 yields

$$Q_2^2 = \frac{R_2}{R_1} (1 + Q_1^2) - 1 \quad (93)$$

For a positive Q_2 , it is thus required that if $R_1 > R_2$, $Q_1 \geq \sqrt{\frac{R_1}{R_2} - 1}$. Similarly, if $R_2 > R_1$, the requirement becomes $Q_2 \geq \sqrt{\frac{R_2}{R_1} - 1}$. Therefore if expressed in terms of Q_L from (90), the condition on which the Π -match is designable is that

$$Q_L \geq \begin{cases} \frac{1}{2} \sqrt{\frac{R_1}{R_2} - 1} & \text{if } R_1 > R_2 \\ \frac{1}{2} \sqrt{\frac{R_2}{R_1} - 1} & \text{if } R_2 > R_1 \end{cases} \quad (94)$$

If the condition in (94) is met, then from (90) and (91), Q_1 and Q_2 can be solved:

$$Q_1 = \frac{2Q_L R_1 - \sqrt{4Q_L^2 R_1 R_2 - (R_1 - R_2)^2}}{R_1 - R_2} \quad (95)$$

$$Q_2 = \frac{2Q_L R_2 - \sqrt{4Q_L^2 R_1 R_2 - (R_1 - R_2)^2}}{R_2 - R_1} \quad (96)$$

Therefore the design procedure of Π -match network is the following:

- 1) With R_1 , R_2 , and Q_L set, check to make sure the condition in (94) is met. If not, reassign their values (usually the Q_L value) until the condition is met.
- 2) Once it is verified that condition in (94) is met, then Q_1 and Q_2 can be calculated from (95) and (96), respectively.
- 3) Solve for B_{C1} , B_{C2} , and X_L from (87) and (92).
- 4) At the frequency of interest, calculate C_1 , C_2 , and L from (86).

Note that by setting one of the capacitances to zero, all the Π -match equations above would become identical to those of the L-match. This confirms that the L-match can be viewed as a special case of the Π -match.

As stated before, for RF power amplifier applications, R_1 and R_2 are usually determined by the power and efficiency optimization and load requirement, respectively. The following discussion thus focuses on

design considerations of Q_L .

First, according to its definition, Q_L determines the bandwidth of the matching network. Therefore, if the bandwidth specification is provided, then

$$Q_L \leq \frac{f_c}{BW} \quad (97)$$

where f_c and BW represent the center frequency and the bandwidth of the band, respectively.

One of the direct tradeoffs in designing Q_L for the pass-band is the requirement of out-of-band harmonic rejections. Less Q_L would provide wider bandwidth, and thus the matching network can be less sensitive to process, voltage, and temperature variations; but high levels of harmonic rejections usually call for higher Q_L values. Basic network analysis can show that the voltage transfer function of the Π -match network is

$$H(s) = \frac{V_2}{V_S} = \frac{R_2}{R_1 + R_2} \frac{1}{1 + s \left[L + (C_1 + C_2) \frac{R_1 R_2}{R_1 + R_2} \right] + s^2 \frac{L}{R_1 + R_2} (C_1 R_1 + C_2 R_2) + s^3 L C_1 C_2 \frac{R_1 R_2}{R_1 + R_2}} \quad (98)$$

Define ω_m as the angular frequency at the match condition, then after some mathematical manipulations of (86), (87), (95), and (96), we have

$$\omega_m C_1 = \frac{2Q_L R_1 - \sqrt{4Q_L^2 R_1 R_2 - (R_1 - R_2)^2}}{R_1 (R_1 - R_2)} \quad (99)$$

$$\omega_m C_2 = \frac{2Q_L R_2 - \sqrt{4Q_L^2 R_1 R_2 - (R_1 - R_2)^2}}{R_2 (R_2 - R_1)} \quad (100)$$

$$\omega_m L = \frac{(R_1 - R_2)^2}{2Q_L (R_1 + R_2) - 2\sqrt{4Q_L^2 R_1 R_2 - (R_1 - R_2)^2}} \quad (101)$$

where $k = R_1/R_2$. Substituting (99), (100), and (101) into (98), the transfer function can be expressed in terms of ω_m , Q_L , and k . In particular, when the harmonic rejection is of concern, the magnitude frequency response is

$$\begin{aligned} |H(j\omega)| = & 2 \left[(k+1) Q_L - \sqrt{4kQ_L^2 - (k-1)^2} \right] / \left\{ \left[2(k+1)^2 Q_L \right. \right. \\ & \left. \left. - 2(k+1) \sqrt{4kQ_L^2 - (k-1)^2} - 2(k-1)^2 Q_L \left(\frac{\omega}{\omega_m} \right)^2 \right]^2 \right. \\ & \left. + \left(\left[3(k-1)^2 - 8kQ_L^2 + 2(k+1) Q_L \sqrt{4kQ_L^2 - (k-1)^2} \right] \left(\frac{\omega}{\omega_m} \right) \right)^2 \right\} \end{aligned}$$

$$+ \left[8kQ_L^2 - 2(k+1)Q_L\sqrt{4kQ_L^2 - (k-1)^2} - (k-1)^2 \right] \left(\frac{\omega}{\omega_m} \right)^3 \right)^2 \Big\}^{1/2} \quad (102)$$

Substituting $\omega = \omega_m$ in (102), the magnitude response at matching condition is

$$|H(j\omega_m)| = \frac{1}{2\sqrt{k}} = \frac{1}{2} \sqrt{\frac{R_2}{R_1}} \quad (103)$$

as expected as the Π -match network can be viewed as an impedance transformer. Define $S(\omega) = |H(j\omega)|/|H(j\omega_m)|$, and once the impedance transformation ratio k is determined, $S(\omega, Q_L, k)$ can be used to determine the lower limit of Q_L if a certain harmonic rejection is specified. For instance, often times the second- and third-order harmonic rejections are specified, and setting $\omega = 2\omega_m$ and $3\omega_m$ in the definition of $S(\omega)$ and substituting in (102) and (103), we have

$$\begin{aligned} S(2\omega_m, Q_L, k) &= 2\sqrt{k} \left[(k+1)Q_L - \sqrt{4kQ_L^2 - (k-1)^2} \right] \\ &\quad / \left\{ \left(\left[(k+1)^2 - 4(k-1)^2 \right] Q_L - (k+1)\sqrt{4kQ_L^2 - (k-1)^2} \right)^2 \right. \\ &\quad \left. + \left[24kQ_L^2 - 6(k+1)Q_L\sqrt{4kQ_L^2 - (k-1)^2} - (k-1)^2 \right]^2 \right\}^{1/2} \end{aligned} \quad (104)$$

$$\begin{aligned} S(3\omega_m, Q_L, k) &= 2\sqrt{k} \left[(k+1)Q_L - \sqrt{4kQ_L^2 - (k-1)^2} \right] \\ &\quad / \left\{ \left(\left[(k+1)^2 - 18(k-1)^2 \right] Q_L - (k+1)\sqrt{4kQ_L^2 - (k-1)^2} \right)^2 \right. \\ &\quad \left. + \left[96kQ_L^2 - 24(k+1)Q_L\sqrt{4kQ_L^2 - (k-1)^2} - 9(k-1)^2 \right]^2 \right\}^{1/2} \end{aligned} \quad (105)$$

Therefore once these harmonic rejection levels are specified, then we can have the second constraint of the Q_L :

$$Q_L \geq \max\{Q_{L,H2}, Q_{L,H3}\} \quad (106)$$

where $Q_{L,H2}$ and $Q_{L,H3}$ are the Q_L that correspond to the specified second- and third-order harmonic rejections obtained from graphical approach according to (104) and (105), respectively.

Another design consideration for specifying Q_L involves the parasitic effect. Assume capacitors are less lossy, and the majority of the loss comes from the parasitic resistance of the inductor, which is usually the case. The Π -match network is redrawn in Fig. 35, with this parasitic resistance r explicitly shown.

The parasitic resistance of the inductor has two effects that would contribute to non-ideal power transfer: a) it causes direct power loss of the network, and b) indirectly, it causes mismatch of the impedance,

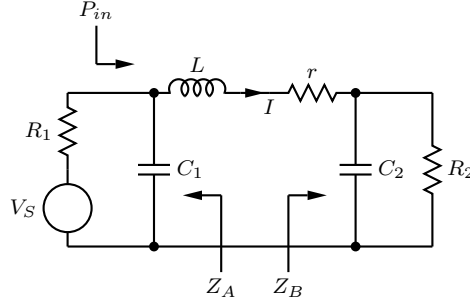


Fig. 35. Π -match network with inductor's parasitic resistance explicitly shown.

thus a portion of the RF power is reflected back and does not reach the output. We denote the network power loss and the reflected power to be P_{loss} and P_{refl} , respectively. Also, define P_{AV} as the available power from the source V_S , P_{in} as the input power at the Π -match network, and P_{out} as the output power dissipated at R_2 . Therefore, due to conservation of energy, we have

$$P_{AV} = P_{in} + P_{refl} \quad (107)$$

$$P_{in} = P_{out} + P_{loss} \quad (108)$$

To analyze the circuit near matching condition, consider the Thevenin equivalent circuit shown in Fig. 36, where V_{Th} represent the Thevenin equivalent voltage, and R_A , R_B , X_A , and X_B are defined in (88).

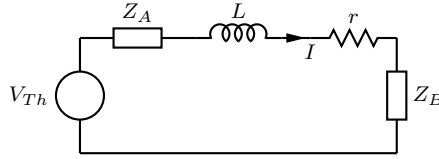


Fig. 36. Thevenin equivalent circuit of the Π -match network near matching frequency.

Also define the “unloaded Q ”, $Q_u = X_L/r$, and since at matching condition, by design, $R_A = R_B$, $X_L = X_A + X_B$, and by definition $Q_L = X_L/(R_A + R_B)$, we have

$$r = \frac{X_L}{Q_u} = 2R_A \frac{Q_L}{Q_u} \quad (109)$$

$$I = \frac{V_{Th}}{R_A + R_B + r + j(X_L - X_A - X_B)} = \frac{V_{Th}}{2R_A} \frac{1}{1 + \frac{Q_L}{Q_u}} \quad (110)$$

By definition [46], the available power can be calculated as $P_{AV} = V_{Th}^2/(4R_A)$. Therefore by using (109) and (110) along with $P_{out} = |I|^2 R_B$ and $P_{loss} = |I|^2 r$, we have

$$\frac{P_{loss}}{P_{AV}} = \frac{2Q_L/Q_u}{(1 + Q_L/Q_u)^2} \quad (111)$$

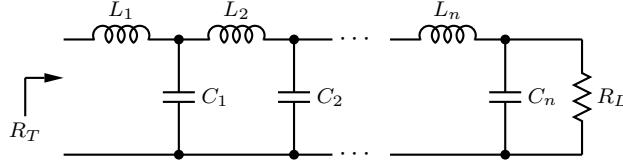


Fig. 37. Multisection matching network.

$$\frac{P_{refl}}{P_{AV}} = \frac{(Q_L/Q_u)^2}{(1 + Q_L/Q_u)^2} \quad (112)$$

$$\frac{P_{out}}{P_{in}} = \frac{1}{1 + 2Q_L/Q_u} \quad (113)$$

$$\frac{P_{out}}{P_{AV}} = \frac{1}{(1 + Q_L/Q_u)^2} \quad (114)$$

Obviously, if the inductor is lossless, i.e. Q_u is infinite, then we have $P_{loss} = P_{refl} = 0$ and $P_{in} = P_{out} = P_{AV}$. If the inductor is implemented off-chip, then usually the Q_u can be in the range of 30 to 80. For power amplifier matching purposes, especially for wireless communication applications, Q_L is usually within 5 to satisfy bandwidth requirement in (97). Therefore, it is a valid assumption that $Q_L/Q_u \ll 1$. Comparing (111) and (112) under this assumption, the direct power loss term, P_{loss} , is the dominant power loss mechanism due to parasitics, and (113) and (114) converge.

If the insertion loss of the matching network is defined as

$$IL = 10 \log \frac{P_{AV}}{P_{out}} \quad (115)$$

then when the available inductor quality factor is given (or its range is known), there can be another design constraint of the Q_L if the minimum insertion loss is specified:

$$Q_L \leq Q_u \left(10^{IL_{min}/20} - 1 \right) \quad (116)$$

In summary, the choice of Q_L can be narrowed down by (97), (106), and (116). Once its value is decided, the design procedure outlined in this section can be used to design the component values of the Π -match network.

V.E. Multisection Matching Network

As the name suggests, multisection matching consists of multiple sections of the basic matching topologies mentioned above. Fig. 37 shows a multisection matching network that is the cascade of L-match sections.

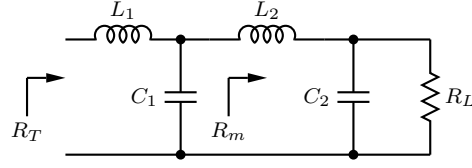


Fig. 38. Schematic of a 2-stage matching network.

Generally speaking, the more number of stages a matching network has, the wider the matching bandwidth [46]. Therefore, for broadband match it is usually desirable to implement multisection match. However, the more components a multisection matching network contains would introduce more power loss. Therefore usually a 2-stage matching network can be used to achieve wider bandwidth, as shown in Fig. 38.

The 2-stage matching network converts the load impedance R_L to an intermediate impedance, R_m , before it is then converted to the termination impedance R_T . Although in theory the choice R_m can be arbitrary, in practice a general fule of thumb is to set R_m such that the impedance transformation ratio of the two stages stays the same:

$$\frac{R_L}{R_m} = \frac{R_m}{R_T} \quad (117)$$

This way the voltage swing progresses evenly through the two stages, resulting in less stress to the components [20].

VI. POWER AMPLIFIER ARCHITECTURES

VI.A. Introduction

As indicated in previous sections, the design of RF power amplifiers faces the tradeoff between linearity and efficiency. Therefore, techniques and architectures to achieve linearity through high-efficient switching PAs and to improve power efficiency of linear PAs have been active research and development topics.

To accommodate more functionalities in battery-powered electronic devices, it is required for the PAs to have high average efficiency. On the other hand, to meet the increasingly stringent linearity requirements in modern communication standards, linearization is often needed even for linear PA classes. In general, architectures and techniques to achieve the above goals fall into two categories: efficiency enhancement techniques and linearization architectures. This section serves as an overview of these previous solutions.

VI.B. Efficiency Enhancement Techniques

Power efficiency would not be quite a big design challenge if the modulation schemes does not involve variations in the envelope amplitude, because in that case a switching mode PA can be used, and the theoretical efficiency should reach 100%. Therefore in frequency modulations (FM) and second-generation (2G) wireless communication standards where the transmission envelope has constant amplitude, much research effort was put to switching PA implementations to achieve power efficiency that were as close to the theoretical value as possible.

To achieve high average power efficiency on the modern electronic devices, it is important to improve efficiency in the power back-off (PBO) region because of the high peak-to-average power ratio (PAPR) modulation schemes used. Therefore, development and implementation of efficiency enhancement techniques have been very active research areas. On one hand, techniques developed a long time ago are implemented using modern technologies, on the other hand, new techniques are developed to improve efficiency.

VI.B.1. Outphasing Modulation: The basic idea of outphasing modulation, as originally proposed by Chireix [51], is to embed the envelope amplitude variations into the phases of two paths, so in each path switching mode PA can be used. The original amplitude modulation would be recovered after a voltage subtractor at the output of the two paths, as shown in Fig. 39.

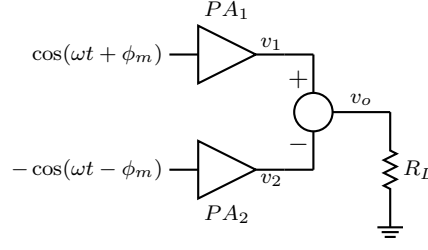


Fig. 39. Conceptual schematic of outphasing modulation.

Suppose the output voltage amplitude of each PA is V , then we have

$$v_1 = V (\cos \phi_m \cos \omega t - \sin \phi_m \sin \omega t) \quad (118a)$$

$$v_2 = V (-\cos \phi_m \cos \omega t - \sin \phi_m \sin \omega t) \quad (118b)$$

Therefore the output voltage is

$$v_{out} = 2V \cos \phi_m \cos \omega t = A(t) \cos \omega t \quad (119)$$

In other words, the amplitude modulation $A(t)$ is embedded into the phase modulations applied to the PA in each path,

$$\phi_m = \cos^{-1} \left[\frac{A(t)}{2V} \right] \quad (120)$$

Although both PAs in an outphasing system can be implemented in switching mode that results in high efficiency themselves, the overall power efficiency still decreases in proportion to the power back-off. Since the DC power is a constant, the shape of the efficiency as a function of PBO is actually the same as that of class A PAs, with the only difference being the maximum efficiency value. There is still one difference between the two systems, in that in an outphasing system, since the PAs are switching mode and high-efficient themselves, they dissipate less heat, therefore outphasing system may have an advantage in reliability performances. In addition, since the power loss in the back-off region is not in the form of heat dissipation, but simply a result of power addition or subtraction, it is possible that this power can be recycled.

To improve power efficiency in the back-off region in outphasing systems, it is necessary to modify the load network. If the PAs are modeled as voltage sources, the load without compensation is shown in Fig. 40. Using the phasor notation, the output voltages of the PAs and the load current can be expressed

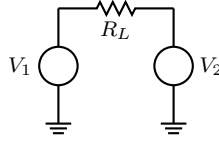


Fig. 40. Demonstration of outphasing load configuration without compensation.

as

$$V_1 = V (\cos \phi_m + j \sin \phi_m) \quad (121a)$$

$$V_2 = V (-\cos \phi_m + j \sin \phi_m) \quad (121b)$$

$$I_L = \frac{V_1 - V_2}{R_L} = \frac{2V \cos \phi_m}{R_L} \quad (121c)$$

Therefore the effective output impedance of the PAs are

$$Z_1 = \frac{V_1}{I_L} = \frac{R_L}{2} (1 + j \tan \phi_m) \quad (122a)$$

$$Z_2 = \frac{V_2}{-I_L} = \frac{R_L}{2} (1 - j \tan \phi_m) \quad (122b)$$

Take PA_1 for example, as a result of outphasing modulation, its output impedance is effectively the half load resistance in series with an inductance whose value depends upon the envelope amplitude. To compensate this, a capacitor can be shunted at its output, and its value is determined by the efficiency improvement required at the certain power back-off. A shunt inductor can be placed at the output of PA_2 accordingly. The resultant schematic is shown in Fig. 41, and the complete analysis is carried out in [51, 52].

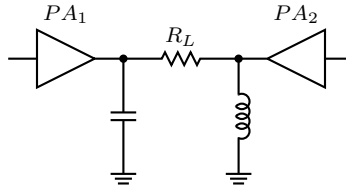


Fig. 41. Outphasing load compensation network.

The circuit shown in Fig. 41 is still not practical since the midpoint of R_L is not a virtual ground. To alleviate this problem, an alternative as shown in Fig. 42 is developed by Raab [53], where quarterwave transmission lines are placed at the outputs of the PAs, and the PA outputs are summed in the current domain. The analysis is similar and not repeated here.

The disadvantage of outphasing technique is that the required digital resource to embed the amplitude

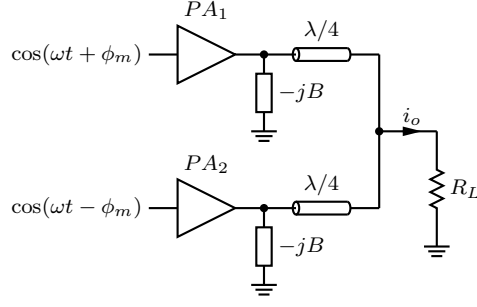


Fig. 42. Outphasing technique implemented in the current domain.

modulation into phase may be large, even in the context of the current DSP capabilities. Also, mismatch between the two paths would result in degradation in both the linearity and system efficiency. Furthermore, the load compensation network component values, as shown in (122), are a function of the embedded phase modulation ϕ_m , but as described in (120), ϕ_m is an inverse trigonometric function of the amplitude modulation $A(t)$, and a relatively large slope of inverse trigonometric functions indicates that the combiner network design could be sensitive to component value variations.

VI.B.2. Doherty Amplifier: Originally proposed by Doherty [54], Doherty amplifier is an architecture that improves efficiency in the power back-off region while maintains the system linearity. In the conceptual schematic shown in Fig. 43, PA_1 and PA_2 are called the carrier and peaking amplifiers, respectively. The purpose of the quarterwave line before the peaking amplifier path is to avoid delay mismatch, whereas the one after the carrier amplifier serves as an impedance inverter. When the power level is low, only the carrier amplifier is activated. Conventionally implemented in class B mode, the carrier PA's output power increases linearly with the input. At a transition point, the carrier PA saturates and the peaking amplifier starts to conduct. As the input power continues to increase, the output current of both PAs increase, but because of the impedance inverter, the output impedance of the carrier PA decreases, and its output power continues to increase while its output voltage stays the same. If the transition point has half the output voltage as that of the maximum output power, then in this medium power region, the carrier PA remains at its maximum efficiency, whereas the efficiency of the peaking PA increases from half of its maximum efficiency at the transition point to maximum at the maximum output power. Therefore the system efficiency would reach maximum both at the maximum output power and at the transition point, and stays relatively high in the medium power region, thus the efficiency at power back-off is improved.

To analyze the Doherty operation, it is important to understand how the quarterwave transmission line can be used as an impedance inverter. Given an arbitrary transmission line with length d and characteristic

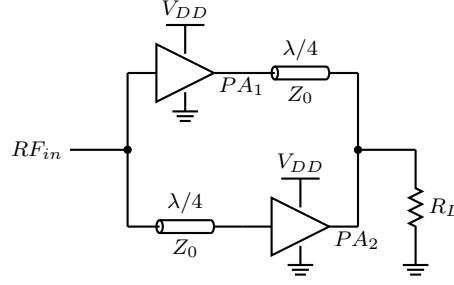


Fig. 43. Conceptual schematic of a Doherty PA.

impedance Z_0 , connected to a load impedance Z_L , the input impedance is given by

$$Z_{in}(d) = Z_0 \frac{Z_L + jZ_0 \tan \beta d}{Z_0 + jZ_L \tan \beta d} \quad (123)$$

where $\beta = 2\pi/\lambda$ and λ is the effective electromagnetic wavelength in the transmission line [46]. If the length of the line is a quarter of the wavelength, then simplification of (123) leads to

$$Z_{in}(\lambda/4) \cdot Z_L = Z_0^2 \quad (124)$$

In other words, the impedance at the two ends of a quarterwave line is inversely proportional to each other, hence the quarterwave line is viewed as an impedance inverter in this regard.

For simplicity, assume the transition point occurs when the output voltage is half of the maximum, i.e. 6 dB back-off from the maximum output power. The analysis in [55] considers a more general case. If both PA are modeled as current sources, then in the medium power region, the Doherty system can be modeled as a circuit shown in Fig. 44.

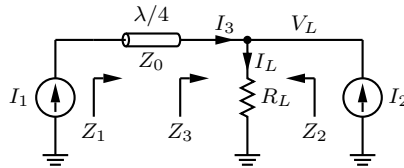


Fig. 44. Simplified Doherty PA for analysis.

Assume that both PAs have the same maximum current I_{max} , and so the RF current amplitude of the PAs at maximum output power, $I_1 = I_2 = I_{max}/2$. In the medium power region, formulate the current from the PAs to be

$$I_1 = \frac{I_{max}}{4} (1 + k) \quad (125a)$$

$$I_2 = \frac{I_{max}}{2} k \quad (125b)$$

where k is a parameter that ranges between 0 and 1, with $k = 0$ corresponding to the transition point and $k = 1$ the maximum output power. If the quarterwave line is lossless, then its input and output voltage and current relationships are

$$V_1 I_1 = V_L I_3 \quad (126a)$$

$$\frac{V_1}{I_1} \frac{V_L}{I_3} = Z_0^2 \quad (126b)$$

Consequently the output current of PA_1 after the quarterwave line

$$I_3 = \frac{V_1}{Z_0} \quad (127)$$

The critical concept of Doherty amplifier can be revealed by calculating the impedances Z_2 and Z_3 :

$$Z_2 = \frac{V_L}{I_2} = \frac{V_L}{I_L} \frac{I_L}{I_2} = R_L \left(1 + \frac{I_3}{I_2} \right) \quad (128a)$$

$$Z_3 = \frac{V_L}{I_3} = R_L \left(1 + \frac{I_2}{I_3} \right) \quad (128b)$$

Thus Z_3 increases as the peaking PA starts to conduct. Due to the impedance inversion by the quarterwave line, Z_1 decreases, hence although V_1 is kept at the saturation voltage level, the output power by the carrier PA continues to increase. Combining the results from (124), (125), and (128b), and solving for V_1 would lead to

$$V_1 = \frac{I_{max}}{4} Z_0 \left[\frac{Z_0}{R_L} + k \left(\frac{Z_0}{R_L} - 2 \right) \right] \quad (129)$$

If $Z_0 = 2R_L$ then according to (129), the output voltage of the carrier PA would be independent of k , and in this case $Z_0 = V_{DD}/(I_{max}/2)$, which is the ideal optimal load impedance of class A or class B PA.

Although the analysis above led to an elegant result both in terms of linearity and back-off power efficiency, Doherty amplifier faces several implementation difficulties. First, the two paths need to have accurate delay match. Superficially this requirement might get translated to matching of the quarterwave lines and the two PAs in each path. However, even if the two PAs deliver the same maximum power and thus have the same dimension, they should be biased differently, and hence would result in different delay because of the bias-dependent parasitic capacitances. In modern wireless communications this difference in delay may be a significant fraction of the RF cycle, and compensation through quarterwave line delays would still make the design sensitive to process, voltage, and temperature variations. Also,

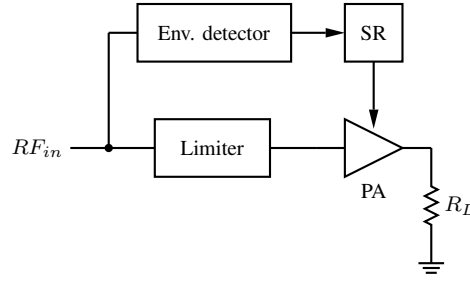


Fig. 45. Conceptual schematic of an EER architecture.

robust implementation of bias circuit for the peaking amplifier is difficult. In addition, as revealed by (129), the characteristic impedance of the quarterwave line should be equal to the optimal load impedance of the carrier PA, which is usually in the order of $10\ \Omega$ or less in low-voltage technologies. Therefore the transmission lines would result in a very wide dimension which would consume a lot of area.

VI.B.3. Envelope Elimination and Restoration Technique: First proposed by Kahn [56], the envelope elimination and restoration (EER) technique separates amplitude modulation from frequency and phase modulations. As shown in a conceptual schematic in Fig. 45, the limiter in the PA path eliminates the envelope amplitude information but preserves phase and frequency modulations. As a result, switching mode PA can be used, yielding a high power efficiency. On the other hand, the envelope amplitude information is extracted by the envelope detector, which in turn controls the switching regulator's output accordingly. This information is restored at the output of the PA because the output amplitude is proportional to the PA power supply, which is the output of the switching regulator.

Modern implementation of the EER system is sometimes termed polar modulation, as shown in Fig. 46 [25]. Advance in digital signal processing (DSP) allows direct generation of amplitude and phase modulations. The phase-locked loop (PLL) in the PA path further reduces phase noise, and the voltage regulator can be implemented by low-drop regulator or switching regulator. Such an architecture provides a possibility of realizing linear transmitters that have high efficiency in a relatively large dynamic range.

Delay mismatch may cause linearity degradations, therefore the system linearity performance is sensitive to process, voltage, and temperature variations unless there is a dynamic feedback correction. Also, large variations on the PA power supply would cause AM-PM conversion and hence phase nonlinearity. In addition, the high PAPR in modern modulation schemes may pose as a design challenge to the voltage regulator, and the trade-off of linearity and efficiency of the regulator may result in an overall efficiency that is less than expected.

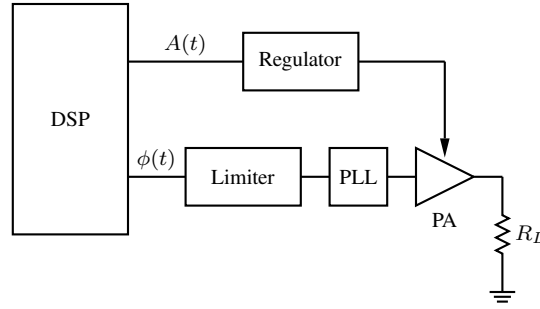


Fig. 46. Conceptual schematic of a polar modulation architecture.

VI.B.4. Envelope Tracking Technique: The conceptual schematic of an envelope tracking (ET) transmitter is shown in Fig. 47. It is very similar with the EER system shown in Fig. 45, but the major distinction is that in an ET system, the PA is usually linear, and thus the variable voltage regulator only needs to provide sufficient voltage headroom for the PA to operate properly in the linear region. Therefore, the amplitude linearity of the system is still provided through the PA, but the design of the regulator can be relaxed since it does not have to accurately follow the envelope amplitude variations.

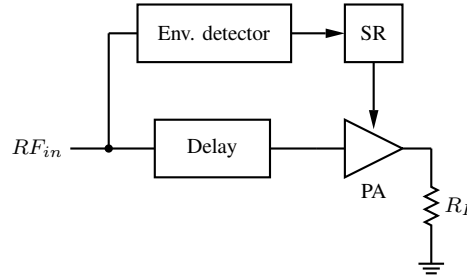


Fig. 47. Conceptual schematic of an ET system.

The drawback of ET architecture is that in low-voltage applications, as the minimum output voltage becomes a significant fraction of the power supply, the efficiency improvement in the power back-off region becomes less. This is especially true in CMOS technologies, thus the ET systems are more common in technologies that has higher breakdown voltages [57–59].

VI.B.5. Power Combining Technique: Power combining technique divides the PA into several sections, and deactivate one or more sections in the power back-off region. As shown in Fig. 48, the output power from individual PA sections are combined in either voltage or current domain [60]. In addition, this technique is a good candidate for multimode applications, where PA sections can be activated or deactivated according to the maximum output power of a certain standard. This technique thus has gained a lot of interest in recent research [60–64].

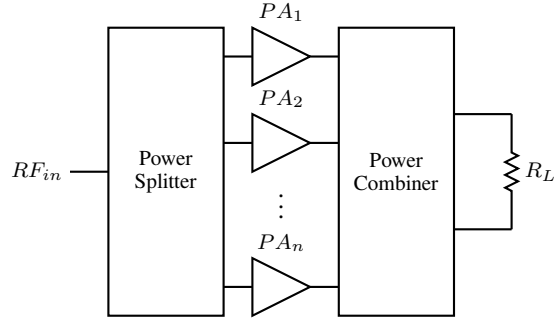


Fig. 48. Conceptual schematic of a power combining architecture.

Switching speed is a design challenge in power combining architectures. Most previous works are able to switch on and off sections of the PA, but not fast enough. Therefore although they can be used for multimode applications and the average efficiency can be improved when part of the PA is deactivated during the low-power mode, the power efficiency *within* each operation mode is the same as that of a stand-alone PA. Another issue with this technique is that the use of on-chip transformers would result in unreliable impedance matching, and as the distance between the metal layers and the silicon substrate continues to shrink, the loss due to substrate coupling would cut into the power saved [65].

This work presents a new method that alleviates the issues above, and the details would be discussed in Section VII.

VI.C. Linearization Techniques

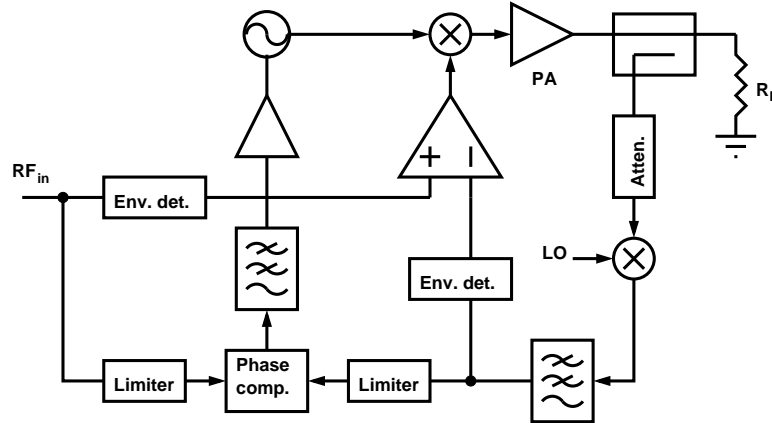


Fig. 49. Conceptual schematic of a polar-loop feedback system.

Because of the overcrowded frequency bands, band-efficient modulation and filtering schemes are widely used in modern communication systems. As a result, the PAPR is high and linearity requirement is

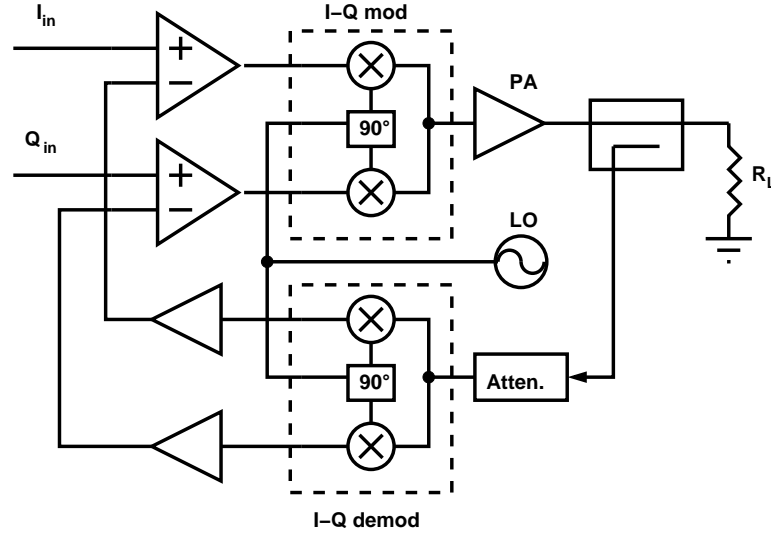


Fig. 50. Conceptual schematic of a cartesian feedback system.

stringent. Consequently, using lineary PA alone may not meet the specific communication standard, due to nonlinearity of the PA itself. Therefore some degrees of linearizations might be needed. In general there are three linearization architectures: feedback, feedforward, and predistortion.

VI.C.1. Feedback Techniques: The system shown in Fig. 49, usually referred to as polar-loop feedback, is a feedback system that consists of two independent loops, one of them corrects amplitude nonlinearity while the other corrects phase distortion. The main design challenge is that loop delay mismatch may affect the system linearity performance. Also, the auxiliary circuits may degrade the system power efficiency.

Cartesian feedback, shown in Fig. 50, alleviates in some degree the delay mismatch problem of polar-loop feedback. However, the complexity due to feedback demodulators and error amplifiers may pose a concern of power efficiency. Another drawback of this system is that it is unable to actively correct AM-PM conversion.

VI.C.2. Feedforward Techniques: The feedforward architecture is illustrated in Fig. 51. Due to the absence of feedback, the bandwidth limitation is alleviated. The error amplifier needs to amplify the error signal, which has a much higher PAPR than the RF signal the PA handles. Therefore the efficiency of the error amplifier may pose as a design challenge. Also, delay mismatch between the two paths might degrade the effectiveness of this approach.

VI.C.3. Predistortion Techniques: Predistortion intentionally introduces nonlinearity to the input signal to cancel distortion from the PA, as illustrated in Fig. 52. Due to advance in the DSP, this technique has gained some popularity. Additional loop can be added for adaptive predistortion, but storage and process

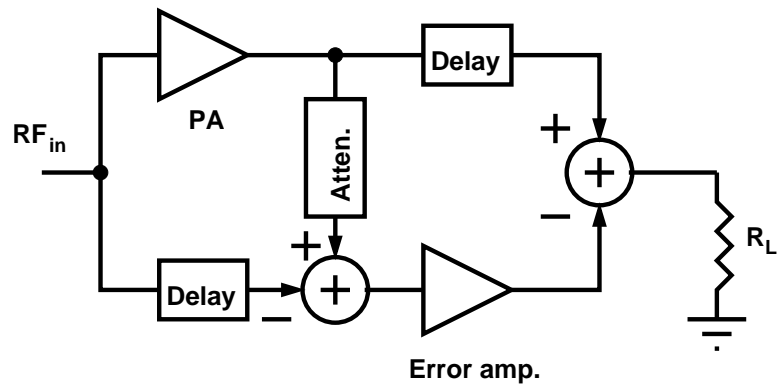


Fig. 51. Conceptual schematic of a feedforward architecture.

overhead, as well as the look-up table updating and convergence time can be an issue [20].

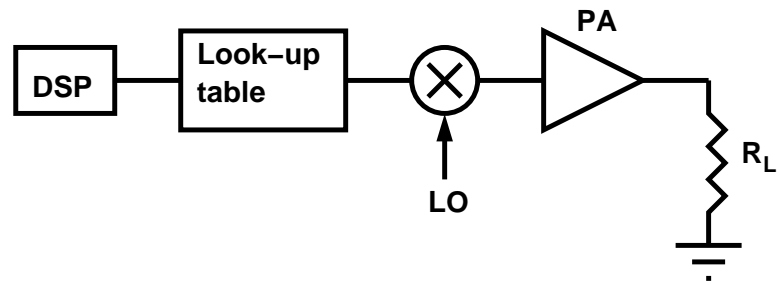


Fig. 52. Conceptual schematic of a predistortion system.

VII. A 35DBM OUTPUT POWER AND 38DB LINEAR GAIN PA WITH 44.9% PEAK PAE AT 1.9GHZ IN 40 NM CMOS¹

VII.A. Introduction

The power amplifier (PA) is one of the major power consumers in the RF transceiver [66–68], and the design and implementation of high efficient CMOS PA has been a very active research and development area during the last few years [69–71]. The 3-5G communication standards use a high data rate and bandwidth efficient modulations that result in a high peak to average power ratio (PAPR). Because of the high PAPR in such modulations during orthogonal frequency-division multiplexing (OFDM), the probability density function (PDF) of the transmitted power will peak in the power back-off (PBO) region. However, the power efficiency of linear PAs reaches maximum at the peak output power, and drops drastically in the PBO region.

Envelope tracking [57–59, 70, 72, 73] and PA segmentation [60–64, 69, 74–76] are two efficiency enhancement techniques that have gained much interest recently. However, the envelope tracking system is becoming less effective in advanced CMOS technologies as the power supply scales down. The minimum drain-source voltage required by PA transistors and the limited drain-source voltage allowed by the technology limit the benefits of this approach; the use of stacked transistors may help to tolerate more signal swing. Additionally, wide bandwidth standards require a high switching frequency switching regulator, which serves as a tradeoff between regulator power efficiency, output ripple, and tracking error [72, 73].

The use of on-chip transformers in segmented PAs usually presents tolerances, especially the magnetic coupling factor that may result in unreliable impedance matching, and as the distance between the metal layers to the silicon substrate continues to shrink, the loss due to substrate coupling would cut into the power saved by such architectures. Therefore, these segmentations must be accompanied by a tunable impedance matching network that makes these solutions sensitive to process-voltage-temperature variations [60, 74]. In this approach, some PA sections are deactivated in the low-power mode, such that overall efficiency for low-power standards is improved. Such architectures do not provide means to improve average efficiency within each mode of operation. On the other hand, the PA based on DAC switching [77] used in polar PAs is an interesting approach that is further exploited in this design.

¹Part of this section is reprinted with permission from “A 35 dBm Output Power and 38 dB Linear Gain PA with 44.9% Peak PAE at 1.9 GHz in 40 nm CMOS”, H. Qian, Q. Liu, J. Silva-Martinez, and S. Hoyos, *IEEE Journal of Solid-State Circuits*, vol. 51, no. 3, March 2016.

In this section, the design of a 1.9 GHz linear segmented PA is presented. To improve efficiency in the PBO region, a combination of PA segmentation and digital signal processing (DSP) is employed. The PA sections are directly connected to the output impedance matching network equipped with class AB common-mode feedback (CMFB) mechanism to reduce common-mode variations when (de)activating the PA segments. The proposed PAs' efficiency in the back-off region is significantly improved since the drivers and PA active sections are correlated with input signal power. The discrete power gain variations were effectively compensated using a digital pre-warping technique employing noiseless, fast, precise, and cheap digital amplification. The digital pre-warping scheme increases the power of weak signals improving the signal-to-noise ratio of the solution under PBO conditions. Preliminary results of this work were recently reported in [43].

This section is organized as follows. Subsection VII.B reviews three popular PA architectures aimed at improving power efficiency in the PBO, namely the envelope tracking system, power combining, and DAC-based technique. In Subsection VII.C, the proposed architecture is described in detail, and an in-depth analysis of the impact of linearity due to timing mismatch is carried out. The design of the PA building blocks is presented in Subsection VII.D, and the measurement results and discussions are presented in Subsection VII.E. Finally, the conclusion is drawn in Subsection VII.F.

VII.B. Efficiency Enhancement Techniques

Current and future generations of communication systems use high PAPR modulation schemes due to the need for bandwidth efficiency and accommodation of multimode and multistandard applications; therefore, the target goal is to improve power efficiency in the PBO region. A brief description of these techniques follows.

a) Envelope Tracking

One of the possible envelope tracking topologies is shown in Fig. 53. Baseband amplitude (the “envelope”) is extracted in the DSP and converted to analog through the digital-to-analog converter (DAC). The envelope signal is then fed into a switching regulator, usually combined with a linear regulator (not shown in this figure) used to reduce V_{DRAIN} ripple. The PA's V_{DRAIN} is dynamically varied, tracking the baseband signal amplitude; PA's efficiency improves at PBO region. One of the major flaws of this architecture is that the timing misalignment of the PA supply voltage to the RF signal will introduce nonlinearity, and most effort to align the two paths are sensitive to process, voltage, and temperature variations. As the CMOS technology scales toward lower breakdown voltages where the PA output voltage

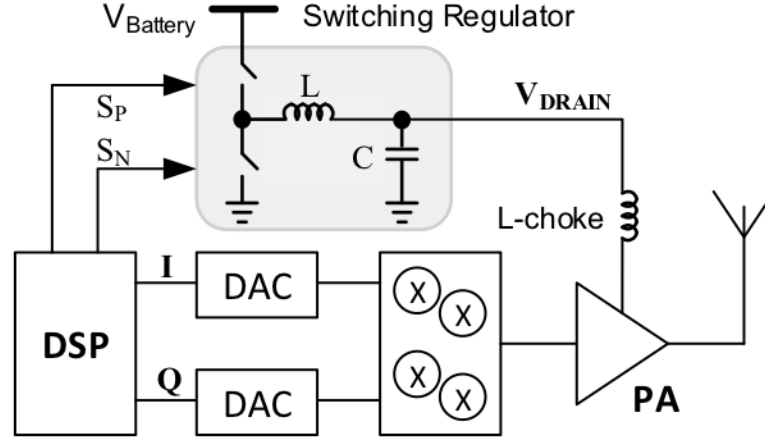


Fig. 53. Conceptual schematic of an ET system.

swing is limited, the envelope tracking technique becomes less effective. On the other hand, the switching regulator must be agile to track the fast variations in the input signal but also with small ripple. These issues demand large switching frequencies, even > 100 MHz for signal bandwidths of 20 MHz with stringent slew-rate specifications. Unfortunately, the increased switching loss of the switching regulator degrades the overall power efficiency when high frequency clocks are employed. Power efficiency degrades due to the use of an auxiliary linear amplifier needed to reduce output voltage's ripple.

b) Power Combining: Segmented PA

The PA can be segmented and the control system deactivates one or more sections depending on the power demanded by different standards, as shown in Fig. 54(a). This approach is well suited for multimode multistandard applications where several sections of the PA can be deactivated when the system is used in low-power mode operation [60–64]. This technique can also be used in switching-mode PAs as demonstrated by [65].

c) Power DAC: Segmented PA

This approach was developed for polar amplifiers; see for instance [77]. It employs a DAC embedded at the output of the RF PA as depicted in Fig. 54(b). The phase of the input signal modulates the carrier and the modulated signal then feeds the linear preamplifiers and so the PA sections $2^M (W/L)_0, 2^{M-1} (W/L)_0, \dots, 2^0 (W/L)_0$. The PA is binary segmented; then, its output current is correlated with the magnitude of the input signal (determined by b_M, b_{M-1}, \dots, b_0) implementing an embedded DAC. In theory, the PA's current efficiency would be maintained close to the maximum attainable in every segment due to the fact that the digital predistortion adjusts the signal power to fit within the

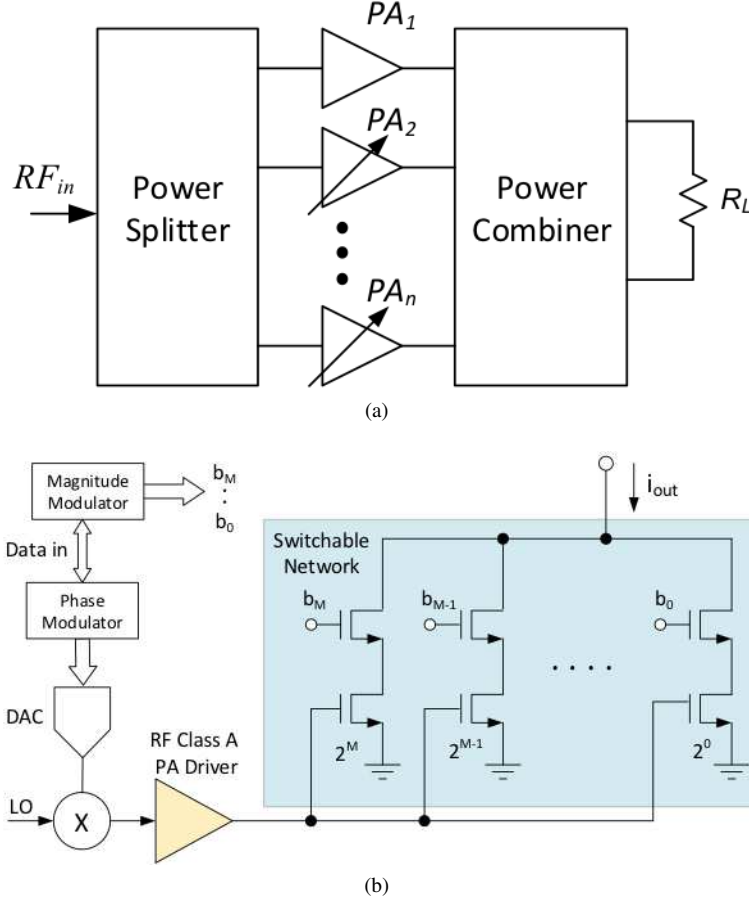


Fig. 54. Conceptual schematic of power combining architecture with switchable PAs. (a) PA for multistandard applications. (b) DAC-based PA with optimized current efficiency.

maximum linear range. However, a number of practical limitations (such as larger dc current than peak ac current in every PA segment is needed for good linearity) degrade it. The PA driver amplifies the phase modulated waveform in a linear fashion to preserve the information, and then demands the use of power hungry class A drivers. Under PBO conditions, the power consumption might be drastically limited by the PA drivers rather than the PA itself. PA drivers can also turn OFF when the corresponding branch is OFF.

VII.C. PA Architecture

Since most communication systems in 3G and onward have a Gaussian distribution power transmission pdf as a function of output power in dBm, the architecture targeted at such communication systems partitions the signal in a linear-to-dB manner to maximize its effectiveness. On the other hand, the best power efficiency in current PAs is obtained for large signals, then the aim of the proposed approach is

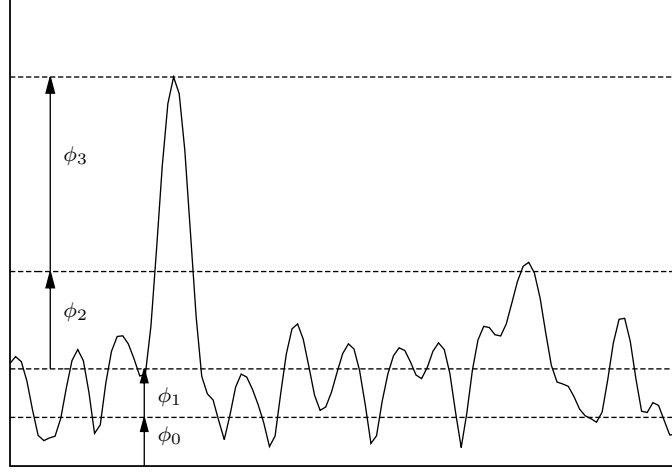


Fig. 55. Correlation between control phases and baseband signal amplitude.

to maintain the PA input signal large; for this purpose, digital prewarping techniques are employed. The incoming signal is segmented into four regions with adjacent regions, which differ in maximum voltage by 6 dB as shown in Fig. 55. More segments can always be used if appropriated for other designs. The four regions are distinguished by the values of the control phases $\phi_1 - \phi_3$. These control bits correspond to the two most significant bits (MSBs) of the baseband signal; thus, baseband signal power is identified in the DSP. The control phases manage the segments of the PA, thus correlating the PA current consumption and gain with the signal MSBs. The prewarped LSBs are then processed using linear amplifiers.

Fig. 56 shows the conceptual schematic of the proposed system. Ignoring the sign bit, the MSBs of the digital representation of the baseband signal magnitude manage the segments, while the least significant are converted into analog format and then up-converted by the mixer. The PA and its driver are divided into four sections in a binary fashion; it is straightforward to realize this operation in the digital domain since the two MSBs provide that information; for better control of the architecture, the MSBs are converted into thermometric format. The control bits $\phi_1 - \phi_3$ drive the PA sections through the drivers. If the signal strength falls in the region ϕ_0 , for instance, the control phases $\phi_1 - \phi_3$ are zero, and then only the unswitchable section manages the signal $S_{in}(t)$. To minimize the switches in the signal path, the drivers are turned OFF by disconnecting the transistor drain from VDD; dc coupling is used to drive the PA sections to avoid the use of large capacitors that introduce significant delay in signal path. The architecture is designed such that when the drivers are turned OFF, the PA sections also shut OFF. As a

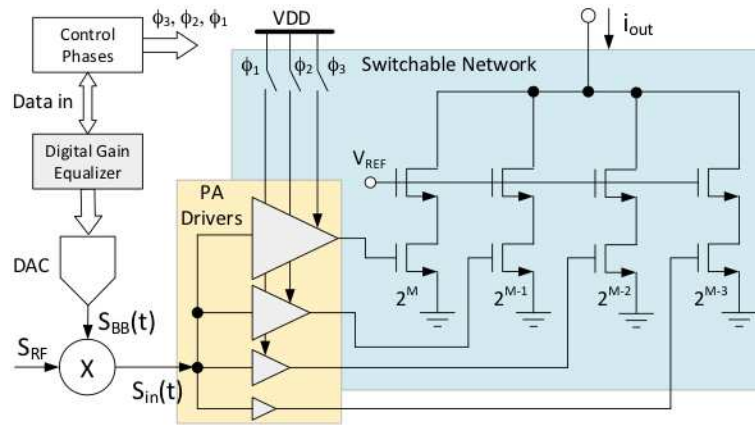


Fig. 56. Simplified schematic of the proposed architecture.

result, the drivers and PA sections are dynamically correlated with signal power providing further power savings.

Due to the manipulation of the segments, the PA power gain follows this pattern, which is a desirable property for polar amplifiers, but makes the PA gain signal dependent for linear amplifiers. An elegant yet efficient solution is to use digital gain equalization to overcome this shortcoming. The signal strength is evaluated and amplified accordingly in the digital domain such that the digital gain and gain attenuation due to PA switching compensate each other leading to a constant power gain factor across all operating conditions. The MSBs used to control the PA segments are also used to manipulate the least significant bits implementing digital gain factors of 2^0 , 2^1 , 2^2 and 2^3 . The realization of these operations is trivial since they correspond to left data shifting by 0, 1, 2, or 3 spots.

A unique property of this approach is that small signals are noise-free amplified in the digital domain, making them more tolerant to thermal noise due to the mixer, PA drivers, and PA sections. The digital amplification does not saturate the RF sections since the magnitude of the prewarped input signal is always within the linear range of the active drivers and PA blocks. The digital gain by multiples of 2 is a very easy and cheap operation since it only requires a bit-shift to the left in the digital domain. If the digital gain equalized signal reaching PA input is fully synchronized with the manipulation of the PA sections, the PA output signal is smooth when transitioning across different segments. However, a common-mode current step (when switching across segments) is an issue that requires further attention.

a) Timing Mismatch Analysis

One concern is the timing alignment of the RF signal path and the digital control phase path. A simplified

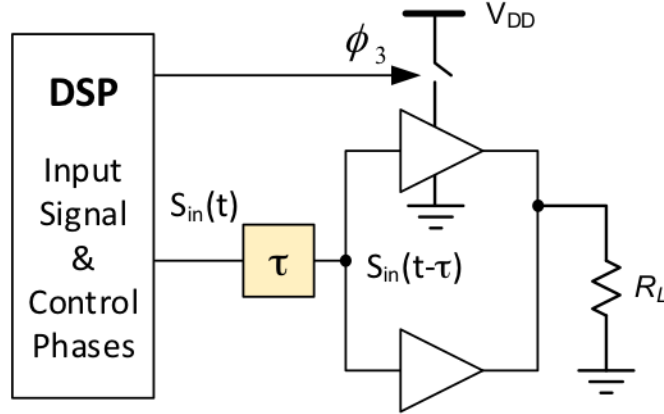
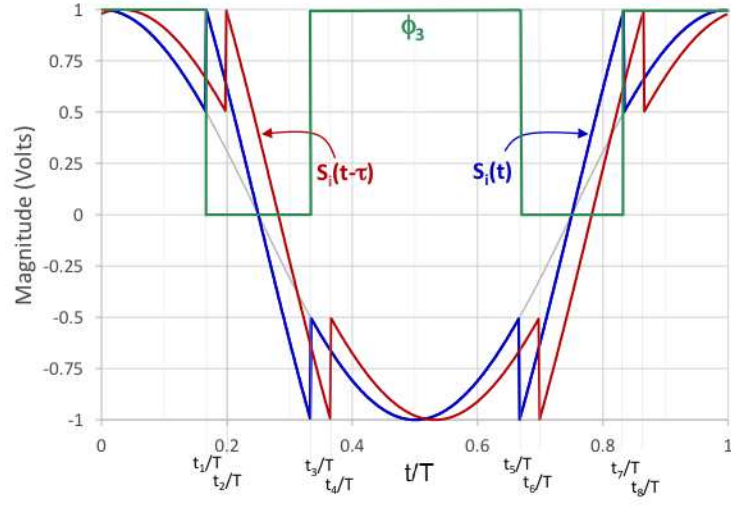


Fig. 57. Simplified model for timing mismatch analysis.

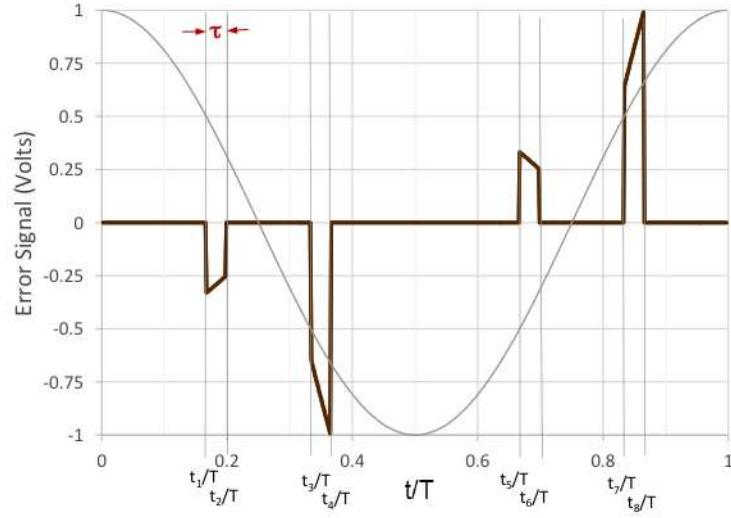
model of the system shown in Fig. 57 is used to capture the essence of the timing mismatch. Let us consider the case of only one-bit control ϕ_3 . Suppose that there is a timing delay of τ seconds between the RF signal path and the control phase, i.e., the control signal arrives at the switch before the corresponding RF signal reaches the PA cells. Assume a modulated input signal $s_{in}(t) = s_{BB}(t)s_{RF}(t) = \cos(\omega_{BB}t)\cos(\omega_{RF}t)$, where ω_{BB} and ω_{RF} represent the baseband and RF angular frequencies, respectively. For simplicity, the amplitude of the input tone and gain of the mixer are chosen to be unity. If all PA sections are active, the output power is then described as $s_{out-N}(t) = 0.5A_{VPA}s_{in}(t)$. However, the baseband equalizer recognizes that the signal power is small and amplifies it by 6 dB; $s_{in}(t)$ is then a pre-equalized version of the original baseband input signal and can be expressed as follows for the case of a single tone:

$$s_{in}(t) = \begin{cases} 2s_{BB}(t)s_{RF}(t) & \text{if } -0.5 \leq s_{BB}(t) \leq 0.5 \\ s_{BB}(t)s_{RF}(t) & \text{if } s_{BB}(t) > 0.5 \text{ or } s_{BB}(t) < -0.5 \end{cases} \quad (130)$$

In Fig. 58, t_i , $i = 1, 3, 5, 7$ which corresponds to the breaking points of the segmentation algorithm. If the timing is perfectly aligned, while the magnitude of the baseband signal is smaller than the threshold voltages, the PA gain reduces by a factor of 2. At the same time, the signal is digitally amplified by two while in this region and, thus, the overall gain remains constant since the digital amplification and PA attenuation are fully synchronized. On the other hand, if there is a timing mismatch of τ seconds between the time we manipulate the PA segments and signal traveling through the up-converter and amplification chain, then the operations are misaligned resulting in an error (glitch like) at the PA output. The delay occurs when the signal travels through the DAC, the mixer, drivers, and PA sections. If the PA sections



(a)



(b)

Fig. 58. PA output waveforms (RF component is not shown for simplicity). (a) Prewarped signal with and without timing delay. (b) Error waveform due to timing mismatch between ϕ_3 and $s_i(t - \tau)$.

are turned OFF earlier, then PA gain drops by 6 dB and stays in this condition until the equalized signal reaches the gate of the PA. This scenario is illustrated in Fig. 58(a) where the PA input signal becomes

$$s_{in}(t) = \begin{cases} s_{BB}(t - \tau)s_{RF}(t), & t < t_2 \\ 2s_{BB}(t - \tau)s_{RF}(t), & t_2 \leq t < t_4 \\ s_{BB}(t - \tau)s_{RF}(t), & t_4 \leq t < t_6 \\ 2s_{BB}(t - \tau)s_{RF}(t), & t_6 \leq t < t_8 \end{cases} \quad (131)$$

where $s_{BB}(t)$ is the incoming baseband signal. Defining the error signal at the output to be the difference between PA output current with timing errors and the ideal output current, then

$$i_e(t) = \begin{cases} -\frac{1}{2}G_{m-VPA}s_{BB}(t)s_{RF}(t), & t_1 \leq t < t_2 \\ G_{m-VPA}s_{BB}(t)s_{RF}(t), & t_3 \leq t < t_4 \\ -\frac{1}{2}G_{m-VPA}s_{BB}(t)s_{RF}(t), & t_5 \leq t < t_6 \\ G_{m-VPA}s_{BB}(t)s_{RF}(t), & t_7 \leq t < t_8 \\ 0, & \text{otherwise} \end{cases} \quad (132)$$

with G_{m-VPA} being the transconductance gain of the PA. The resulting error signal is plotted in Fig. 58(b); the RF component is not shown to simplify the plot. In general, the error signal resulting from the timing mismatch would be manifested as the convolution of the signal LSBs with a time delay of τ seconds, the MSBs, and a time window of τ seconds. For the sake of simplicity, let us denote $\theta = \omega_{BB}t$; then, the third Fourier coefficient of the error signal can be calculated as follows:

$$a_3 = \left(\frac{G_{m-PA}}{\pi} \right) \left(-\frac{1}{2} \int_{\theta_{1,5}}^{\theta_{1,5}+\theta_\tau} \cos \theta \cos 3\theta d\theta + \int_{\theta_{3,7}}^{\theta_{3,7}+\theta_\tau} \cos \theta \cos 3\theta d\theta \right) \quad (133)$$

where $\theta_i = \omega_{BB}t_i$, $i = 1, 2, 3, 4$ and $\theta_\tau = \omega_{BB}\tau$. Calculating of the integrations and then rearranging the expanded terms, noting from Fig. 58 that $\theta_1 = \frac{\pi}{3}$, $\theta_3 = \frac{2\pi}{3}$, $\theta_5 = \frac{4\pi}{3}$, $\theta_7 = \frac{5\pi}{3}$, would lead to

$$a_3 = \left(\frac{\sqrt{7}G_{m-PA}}{2\pi} \right) \left[\frac{1}{2} \sin 2\theta_\tau \sin (2\theta_\tau - \phi) - \sin \theta_\tau \sin (\theta_\tau + \phi) \right] \quad (134)$$

where $\phi = \tan^{-1} \left(\frac{1}{3\sqrt{3}} \right) = 0.19 \text{ rad}$. If we assume that $\theta_\tau \ll \phi$, then (134) reduces to the simpler yet intuitive result

$$|a_3| \approx \left(\frac{\tau}{T_{BB}} \right) G_{m-PA} \quad (135)$$

where T_{BB} is the baseband signal period. Since $a_1 \approx G_{m-PA}s_{BB-pk}$ in this simplified analysis, the

third-order intermodulation distortion due to the timing mismatch IMD_3 is proportional to $\frac{3\tau}{4T_{BB}}$. For a baseband signal of 10 MHz ($T_{BB} = 10^{-7}$ s), the delay error τ must be under 1.3×10^{-9} secs to maintain IM_3 under -40 dB. Timing mismatches in other 3 PA segments add similar effects and increase the PA sensitivity to time delay mismatches. Even more, in practice the computation is more complicated since the spectral leakage is the result of the convolution of the MSBs used to control $\phi_1 - \phi_3$ and the signal power of the least significant bits $s_i(t)$ and a time window of τ seconds correlated with the MSBs; notice in (132) and (133) that the magnitude and sign of the windowing is function of the direction of the transition of the MSBs: -1/2 when the MSB transition from 1 to 0 and +1 when transitioning from 0 to 1.

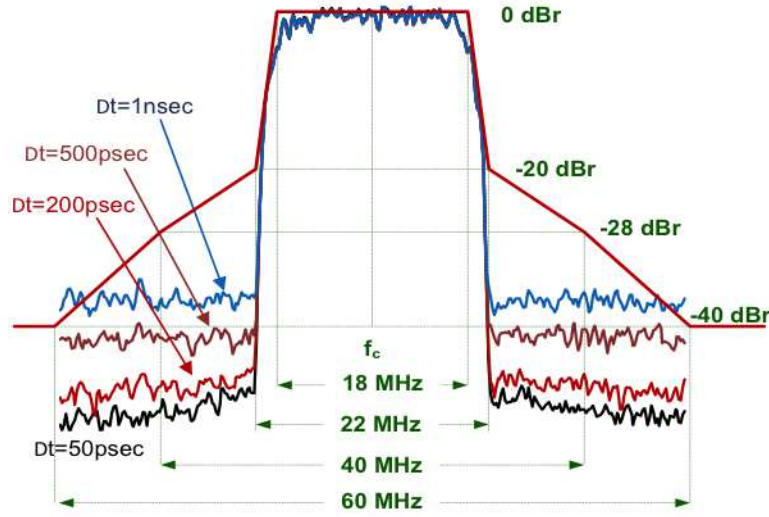


Fig. 59. Timing mismatch effects on ACLR.

To reduce the nonlinearity caused by the timing mismatch, a delay cell is added to the system to reduce the timing mismatch, as shown in Fig. 57. The delay cell includes a replica of the preamp, but it acts as a digital driver. After fine tuning the size of the delay cell using extensive post-layout simulations, the on-chip delay mismatch was under 100 ps for all segments and under PVT variations. Timing mismatches generate glitches (MSBs and least significant bits are not well aligned as depicted in Fig. 58) that may not significantly degrade the received constellation if properly sampled at the receiver. These effects have more effect on ACLR since these glitches are signal dependent. Extensive simulations in a WCDMA system, where the channel bandwidth is 3.84 MHz, a timing mismatch of 500 ps would result in PA neighbor channel leakage power under -40 dB as illustrated in Fig. 59. The timing delay block was not manipulated during characterizations.

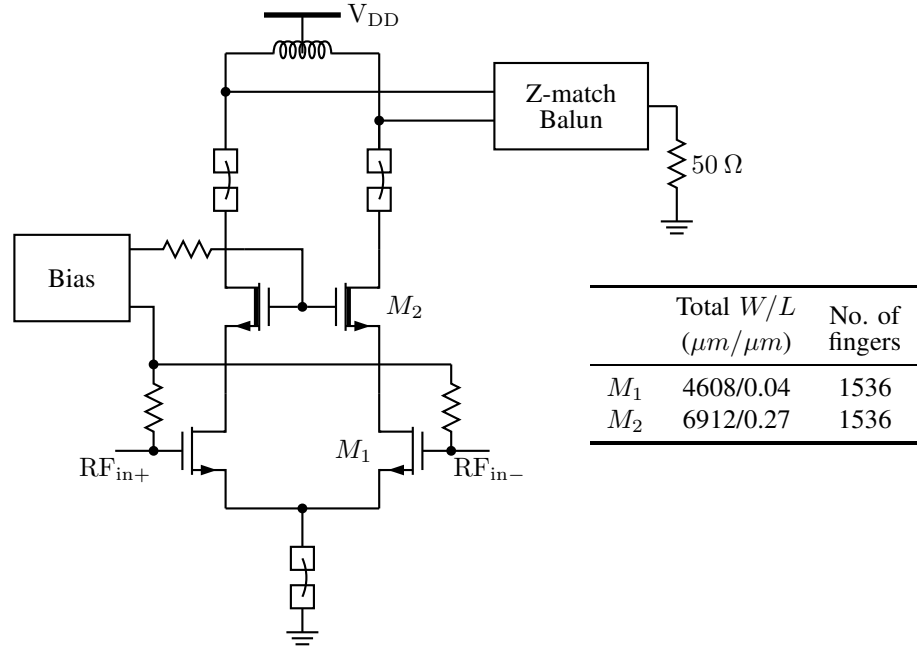


Fig. 60. Schematic of the PA output stage; the core consists of 1536 replicas.

VII.D. PA System Design

The critical design of the proposed system is the switching scheme, which is applied to both the PA sections and their drivers. The PA design details are described in this subsection.

a) Output Stage Design

Fig. 60 shows the schematic of the PA stage. Cascode configuration is used to improve its reliability. The common-source transistors are standard thin oxide transistors that have lower input capacitance and higher transconductance; the common-gate transistors have thick oxide to withstand larger voltage swing. At maximum RF output power, the voltage swing at the drain terminal of the cascode device and the common-source device are 2.5 and 0.75 V_{pk} , respectively. The transistors are optimized for linearity, and their sizes are also included in Fig. 60. The nominal gate overdrive voltage for the transistors are $V_{OV1} = 300$ mV and $V_{OV2} = 400$ mV. At maximum RF output power, the simulated bias current of the output stage and driver stage are 980 and 320 mA, respectively. Maximum RF current is expected at this stage; thus, extra care is needed in the design layout. Multiple pads for the output and ground nodes are used, and the ground pads of this stage are not shared with the remaining parts of the chip. The bondwires are explicitly drawn to indicate that those pads are for the output stage exclusively. The transistors are organized in clusters employing common-centroid techniques to facilitate the connectivity and to minimize transistor mismatches. The PA transistors are dc connected to the PA drivers; thus, no

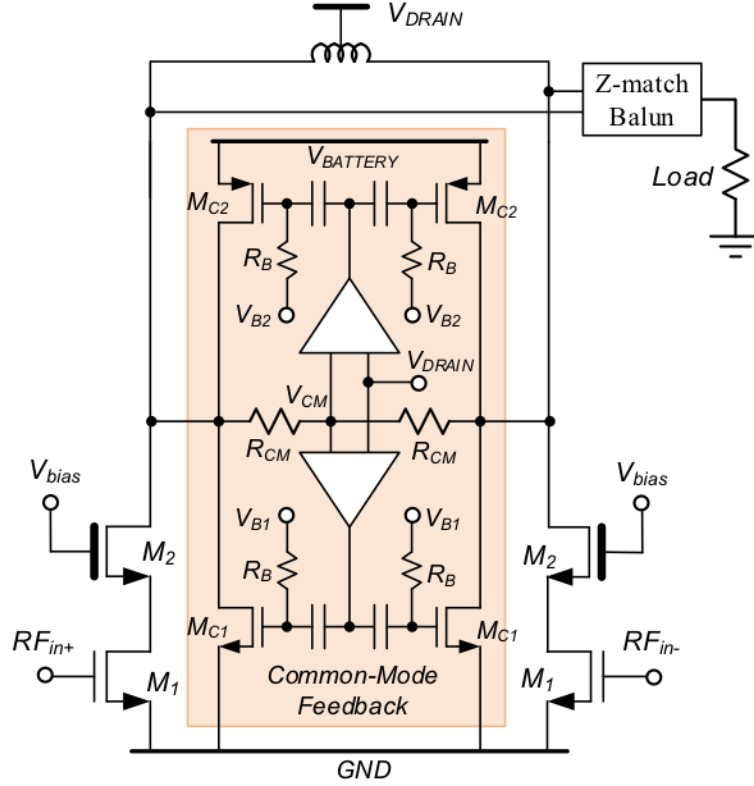


Fig. 61. Schematic showing the CMFB circuit allocated at PA output.

additional switches are required to enable or disable these sections. When M1 transistors are switched ON/OFF, there is a significant common-mode step in current that may produce significant common-mode ringing and up to 1 V common-mode peak variation. To alleviate this issue, a fast class AB CMFB circuit shown in Fig. 61 is allocated at PA output. A couple of single stage amplifiers compare the common-mode output signal and V_{DRAIN} and drive the class AB amplifiers composed by transistors M_{C1} and M_{C2} . These transistors are biased through R_B and $V_{B1,2}$ at the onset of subthreshold region to save power. Class AB amplifiers M_{C1} and M_{C2} minimize the power consumption but are able to deliver/sink enough instantaneous current reducing the common-mode glitches generated by the transistor's switching.

b) PA Drivers

The schematic of the driver stage is shown in Fig. 62. It consists of a differential pair with resistive load, a switch controlled by the control code, and a CMFB loop. C_p and C_{PA} in Fig. 62 represent the effective parasitic capacitance at the common-source node and the output nodes, respectively. Direct coupling between the driver and the PA stages reduces the switching time. When the switch is opened, the

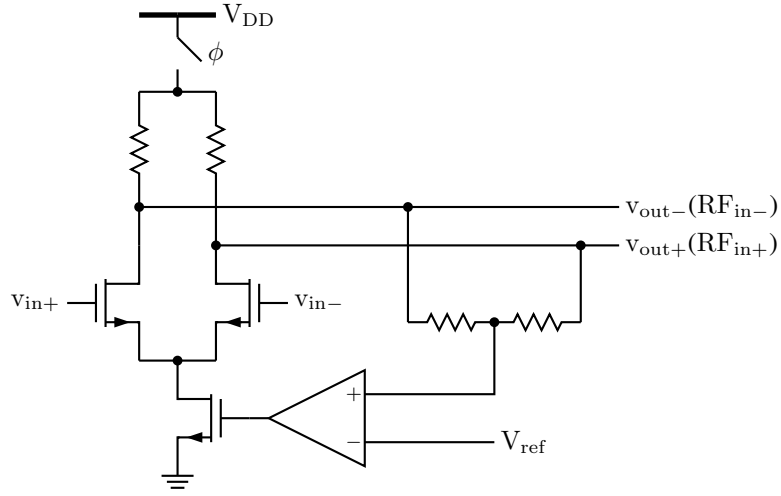


Fig. 62. Conceptual schematic of the driver stage.

driver's output common-mode voltage moves down very quickly, putting the differential pair transistors in the triode region. The common-mode voltage drops, and then breaks the loop during this condition, which helps turning down quickly the preamplifier outputs. When the switch is closed again, the output voltage of the driver moves toward V_{DD} and is only limited by the time constant $R_L C_{PA}$. Since the load resistors R_L are small (in this case around $30\ \Omega$ for the unit-cell driver), the time constant is small, and fast low-to-high transition is achieved. As soon as the common-mode level exceeds the reference voltage, the loop tries to reach its steady state; then, settling time of the CMFB is function of the loop properties. Therefore, the use of fast CMFB is a must.

Open-loop gain, closed-loop bandwidth, and stability are all important parameters to be considered when designing the CMFB loop. Simulation results of the common-mode voltage as the switching takes action are illustrated in Fig. 63. The common-mode voltage moves very quickly until 400 mV is reached because the loop is still broken due to the lack of current in M_4 . The knee during the rising transition is due to the fact that the fast voltage variation at the drain of M_3 put them in a saturation region, allowing the generation of instantaneous discharging current until the parasitic capacitor C_p can get charged. Then, the drain current of M_3 reduces again and the common-mode voltage rises very fast again until reaching its steady-state condition. The 1% settling time under the worst-process corner is less than 8 ns, which means that even if the baseband signal bandwidth is 10 MHz, the switching process would only take 8% of the signal time period in the worst case. The common-mode settling time issues arise during the transition time of the incoming data, generating data-dependent glitches that may degrade the ACLR and EVM figures.

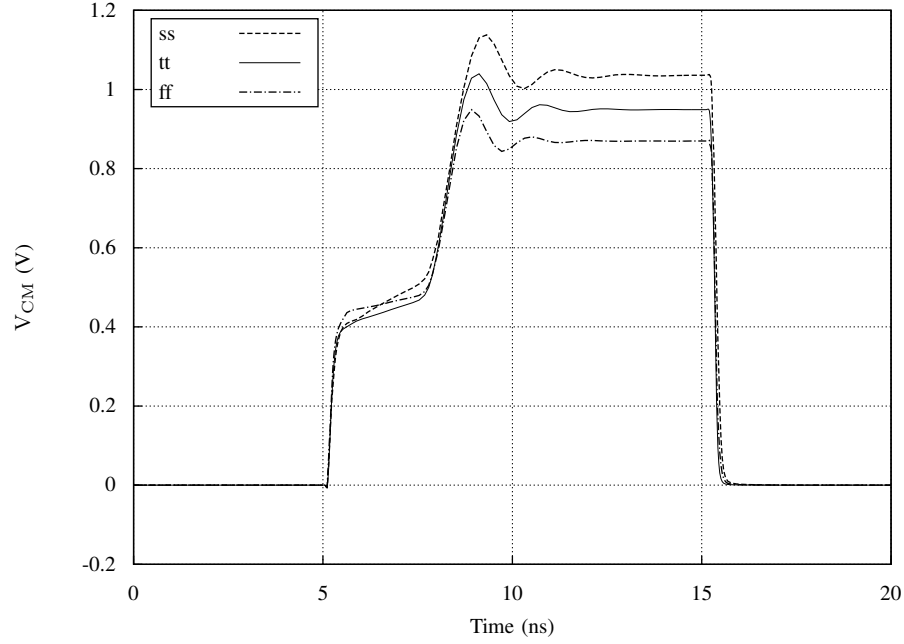


Fig. 63. Simulation results of common-mode voltage transient response.

When all the PA and driver sections are active, the simulated power gain of the driver stage is about 18 dB, whereas the power gain of the output stage is around 20 dB.

c) Output Impedance Matching Design

For the output impedance matching circuit, a multisection network was implemented. Fig. 64 shows a half representation of the matching network. C_D and L_{bnd} stand for the drain capacitance and the bondwire inductance, respectively. The parasitic capacitance at the package on the PCB is accounted for in C_1 . The transmission line with a length d and characteristic impedance Z_0 is formed by a microstrip line consisting of the PCB trace and the ground plane underneath. R_L represents the input impedance of a balun, which is $25\ \Omega$ for a half circuit. The optimal PA load impedance R_T is determined by maximum linear output power design specification. The loadpull simulations further allow us to optimize the choice of R_T . The RF choke inductor is off-chip and not shown in Fig. 64. Its value is chosen such that at RF, it is seen as a high impedance to the PA, while at the switching frequency of the switching regulator, it is seen as a low impedance to the switching regulator. The matching network component values can be determined by hand calculations, the Smith chart, or existing software packages. The summary of the component values is given in Table VIII.

Two important design specifications for the output matching network are bandwidth and insertion loss.

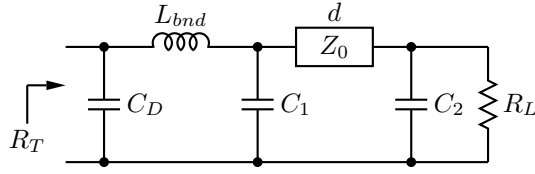


Fig. 64. Two-section impedance matching network.

TABLE VIII
Impedance Matching Network Component Values

C_D	C_1	C_2	L_{bnd}	Trans. Line	
				W (mil)	d (mil)
10	15.9	3.27	400	150	244

The matching circuit used in the proposed system is effectively a multisection design, and its bandwidth is sufficient for WCDMA applications. To ensure robustness, the insertion loss of the output matching is simulated under process variations, as shown in Fig. 65. The worst case of the simulated insertion loss is around 1 dB when all component values are shrunk by 30%. However, this is the less likely case, since nonidealities usually result in additional parasitic components, making the effective component values larger. If all component values increase by 30%, the insertion loss is simulated to be only 0.2 dB.

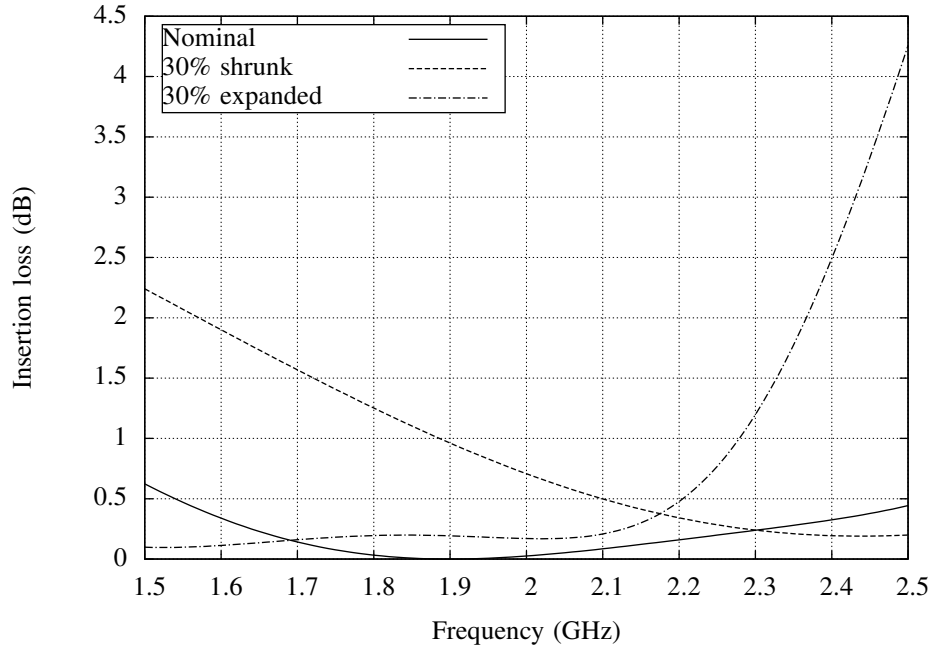


Fig. 65. Insertion loss simulation with process variations.

The Q of the bondwire inductance was assumed to be 50 in simulations. Both the change in Q and

the value of the bondwire inductance affect the output matching network. This effect manifests itself in higher insertion loss at frequency of interest. The insertion loss at 1.9 GHz was simulated with various Q and L values of the bondwire inductance across all four modes of operations. Under extreme conditions, i.e., inductance increased by 25% and $Q = 30$, the insertion loss barely exceeded 1 dB.

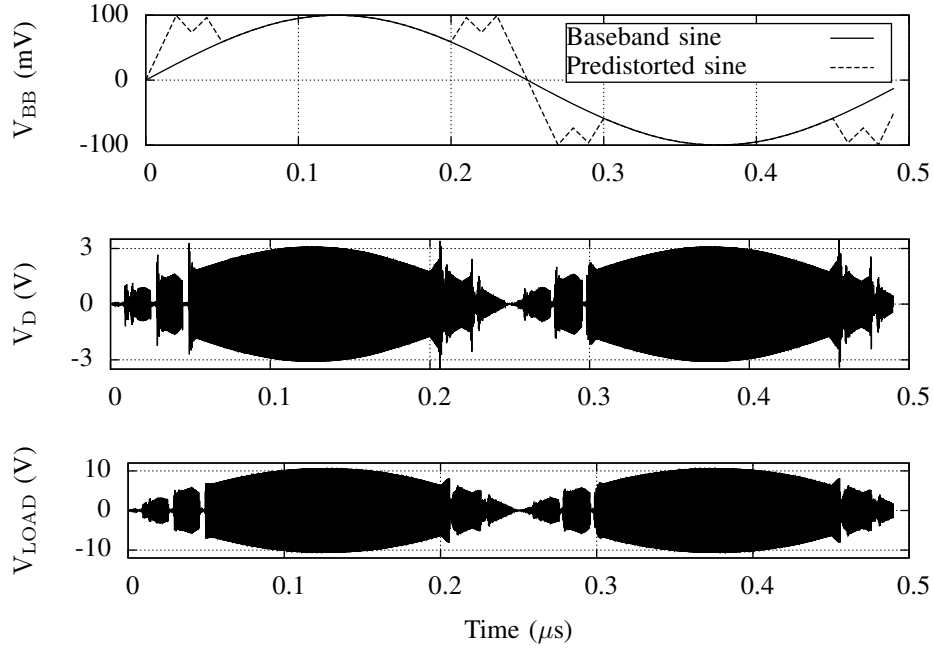


Fig. 66. Transient simulation results. (a) Input signal before and after digital prewarping (top trace). (b) Output signal at drain voltage (middle trace). (c) Output signal after impedance matching network.

As the PA switches ON and OFF different transistor sections, the output impedance of the transistor changes. However, the transistor is in either an active region or a cutoff region, and the drain capacitance is mainly due to the depletion region capacitance between the drain and the substrated plus the gate-drain overlap capacitance. To test if the impedance matching network works properly in each switching scenario, a 2 MHz sinusoidal baseband signal modulated to the carrier frequency is applied to the system. The transient waveforms are shown in Fig. 66. The top plot shows the original sinusoidal baseband signal, along with a predistorted baseband that is to be input to the PA. As shown in Fig. 66, the impedance matching network functions properly at all switching scenarios. The peak differential voltage amplitude before and after the impedance matching network are 3.1 and 10.7 V, respectively. The voltage transformation ratio of 3.45 thus implies an impedance transformation ratio of 11.9, as desired ($Z_L/Z_T = 50/4.5 = 11.1$). Notice that if the mismatches in segmented PA are small, the ac current delivered to the matching network is smooth for the entire power range. In practice, some glitches are present when switching between segments

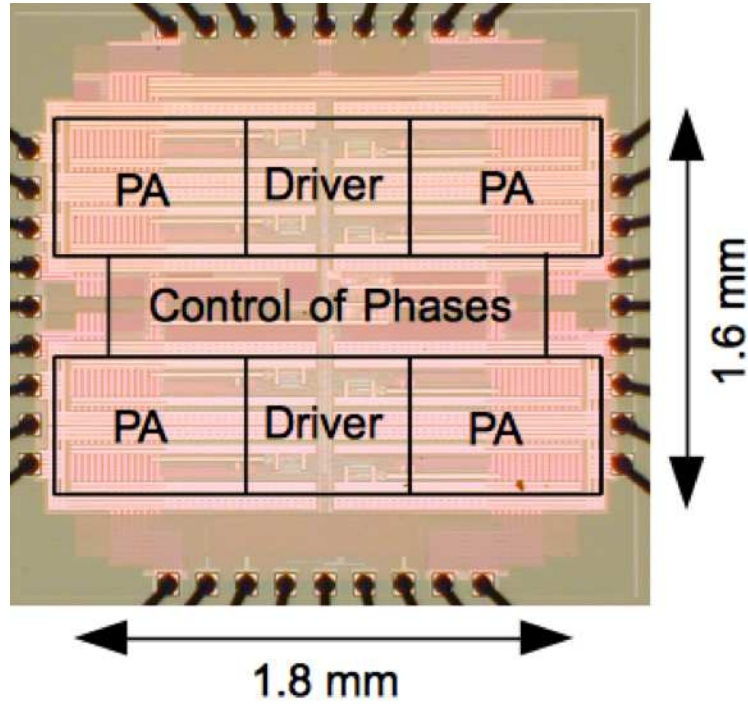


Fig. 67. Microphotograph of the chip.

mainly due to the unavoidable parasitic capacitors and timing offsets.

VII.E. Measurement Results

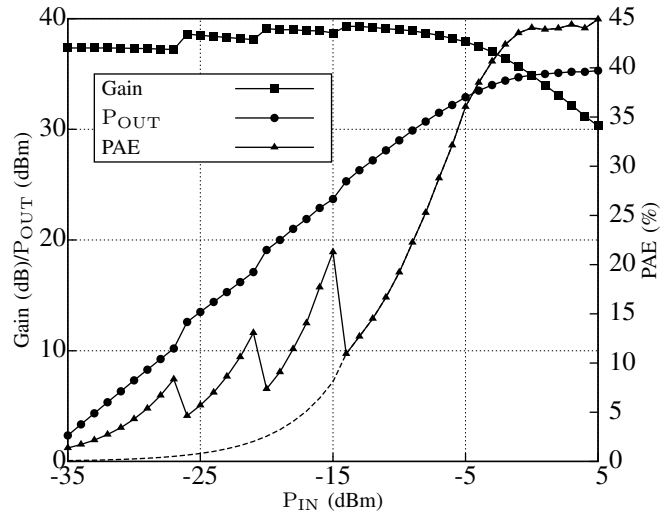


Fig. 68. Measured gain, output power, and PAE as a function of input at 1.9 GHz.

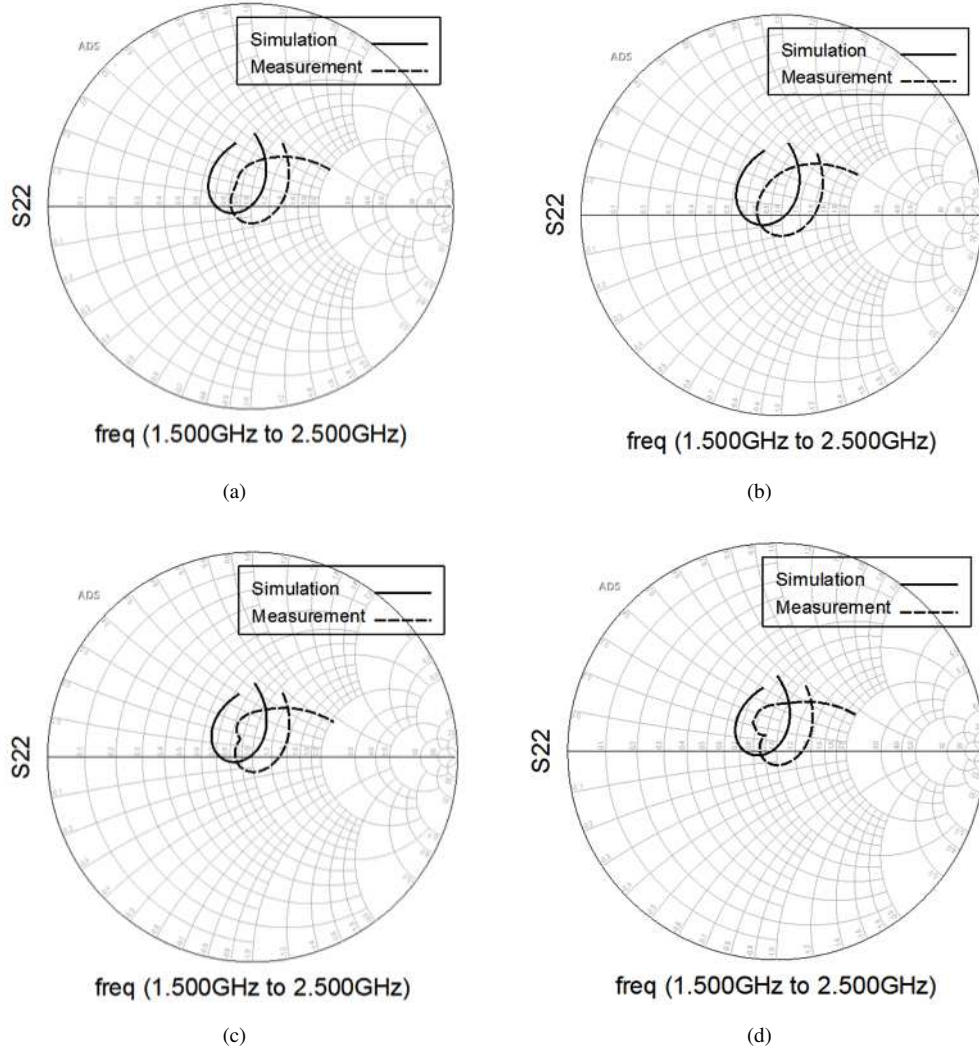


Fig. 69. Simulation and measurement results of PA's S22 when the control bits are (a) 111; (b) 011; (c) 001; and (d) 000.

The PA was fabricated in a TSMC 40 nm CMOS process, and Fig. 67 shows the microphotograph of the chip. The chip area is approximately 2.88 mm². A single-tone continuous-wave (CW) signal of 1.9 GHz was applied to characterize the PA in all four operation modes. Fig. 68 shows the measured gain, output power, and power-added efficiency (PAE) as a function of the input power. The PCB and cable losses are de-embedded in the performance. The PA's output P_{1dB} and P_{SAT} are measured as 31 and 35 dBm, respectively. The average power gain is 38 dB, and the PAEs at P_{1dB} and P_{SAT} are 28.8% and 44.9%, respectively. As a comparison, the PAE of the PA without the proposed power efficiency improvement techniques was measured, and displayed in Fig. 68 as the dashed curve. The PAE improvement in the PBO region is apparent. For instance, at 20 dB back-off from P_{SAT} , PAEs of the PA with and without

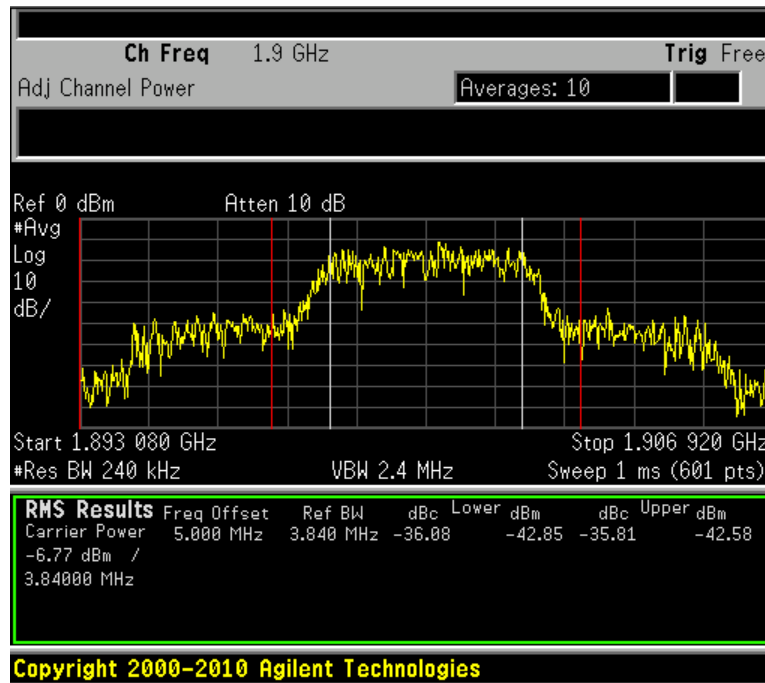


Fig. 70. ACLR measured at maximum output power of 31 dBm.

segmentation are 21.3% and 8.1%, respectively. If required, more segments can be added to improve PA power efficiency at higher power levels.

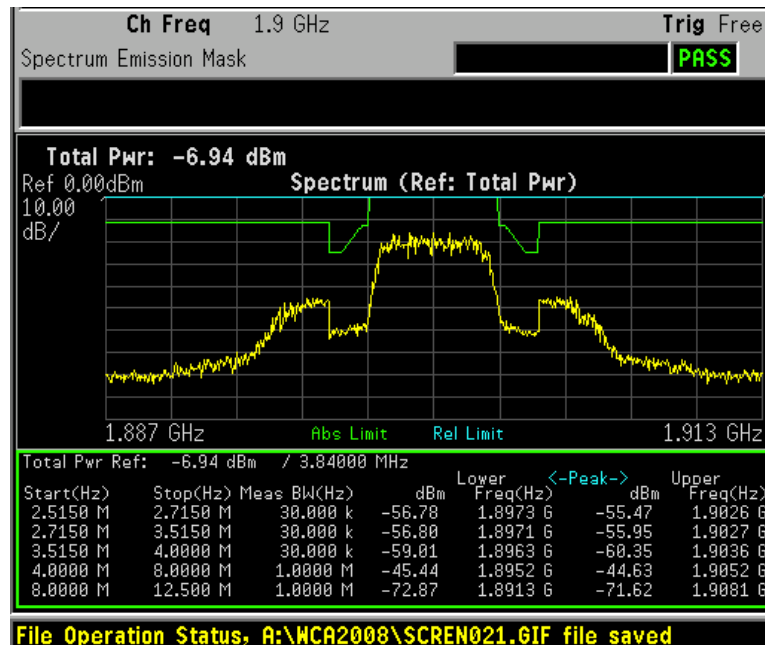


Fig. 71. SEM measured at maximum output power of 31 dBm.

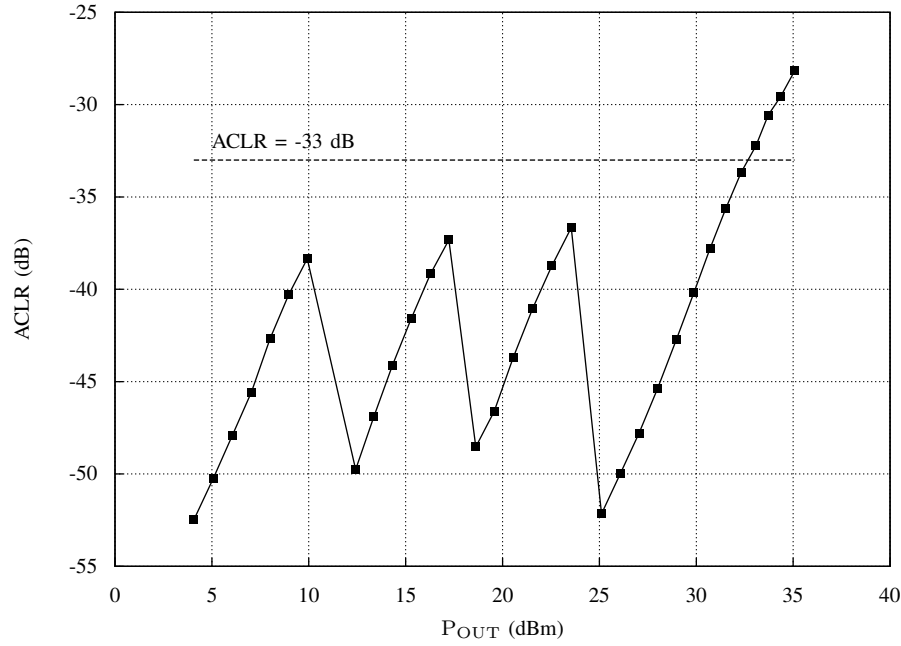


Fig. 72. ACLR as a function of maximum output power.

The S_{22} of the PA in each mode of operation is simulated and measured as shown in Fig. 69. Although there is some mismatch due to nonidealities, it is manageable, and can be optimized by tweaking the output matching network component values. Note that both simulation and measurement show that S_{22} does not vary much across different modes of PA operation. due to the cascode topology of the output stage, PA's output resistance is kept large as compared with the transformed R_T ; therefore, the variation in PA output resistance is absorbed by the output matching network.

A WCDMA baseband signal that is compliant with the 3GPP standard [31] was generated, preprocessed in digital, and up-converted to 1.9 GHz with a bandwidth of 3.84 MHz. According to [31], the adjacent channel leakage ration (ACLR) at ± 5 MHz should be kept below -33 dBc for cellular handsets to comply with the standards. The measured output power spectrum is shown in Fig. 70, with the PA under test transmitting a maximum linear power of 31 dBm. Data analysis from the spectrum analyzer shows that the ACLR at a maximum power of 31 dBm is -35.8 dBc. The spectrum emission mask (SEM) measurement was carried out, and the result is shown in Fig. 71. Under maximum output power condition, the PA meets the 3GPP SEM specifications.

With a fixed set of switching thresholds for the PA and the switching regulator, the ACLR as a function of maximum output power was measured, and the result is shown in Fig. 72. Contrary to classic PA

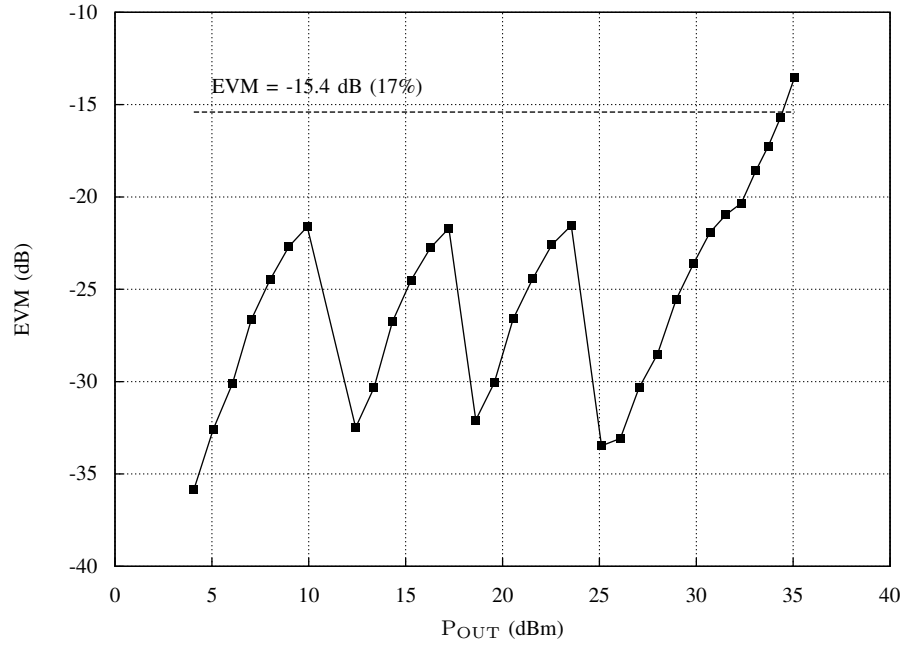


Fig. 73. EVM as a function of maximum output power.

cases where amplifier's linearity improves at low power, the linearity of the proposed architecture was compromised at the PBO region due to the digital amplification. The PA, however, still met the required specifications. This is because the PA transistors work close to their maximum power capacity most of the time. These results show that the proposed architecture has a good balance between power efficiency and linearity. Another linearity figure of merit is the error vector magnitude (EVM). For 3G WCDMA, the specification for EVM is less than -15 dB (17%). The EVM as a function of the maximum output power of the PA is shown in Fig. 73. At a maximum output power of 31 dBm, the EVM is -21 dB (8.9%).

The phase error is measured as an indication of the PA's AM/PM nonlinearity. The phase error as a function of the output power is shown in Fig. 74. The phase error is under 2.5% up to 35 dBm output power.

Recently reported linear PAs with segmentation technique to improve PAE were compared with the proposed PA in Table IX. The proposed PA achieved a remarkable peak PAE as well as outstanding P_{SAT} and marks at PBO regions. Such an improvement was achieved by the combination of segmentation and proposed digital predistortion technique. Moreover, the proposed PA enables switching between different modes within a very short time frame, which is the first to report such a feature, to the author's best knowledge.

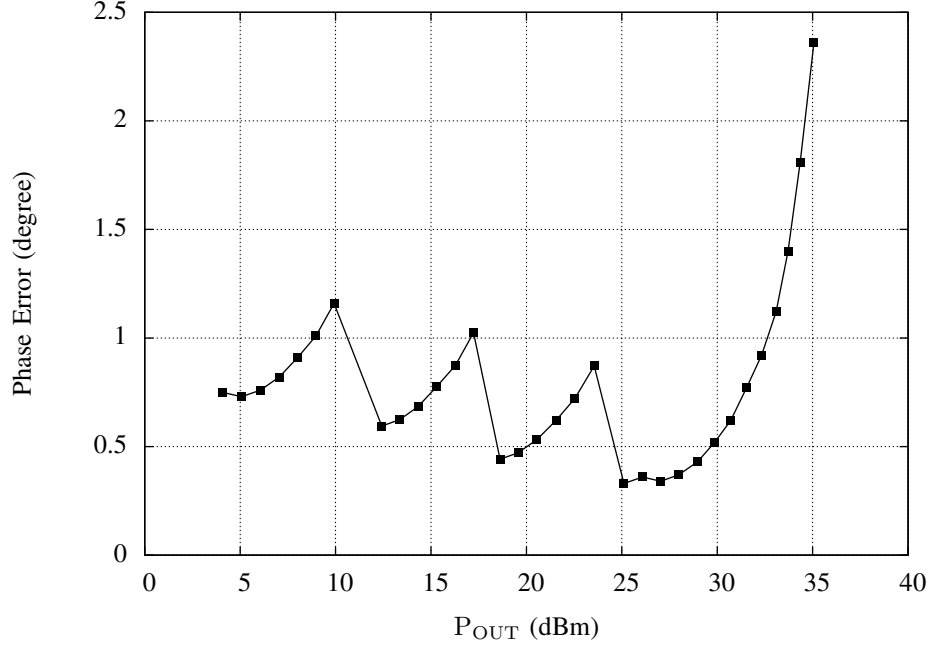


Fig. 74. Phase error as a function of maximum output power.

TABLE IX
Comparison With Recently Published Works

Reference	Frequency (GHz)	P_{SAT}/PAE (dBm/%)	V_{DD} (V)	CMOS (nm)	Size (mm ²)	Number of Modes	PAE Increase at PBO(%)		
							7 dB	10 dB	15 dB
[62]	2.4	23.1/42	1.5	130	5.48	2	3.3	4.3	3.6
[63]	2.4	27/32	1.2	130	2	2	5.4	4.0	3.4
[64]	2.4	23.1/42	3.3	180	0.88	2	8.7	8.7	7.2
[75]	2	23/38	2.5	250	2.48	3	10	N/A	N/A
[76]	2.45	31.5/25	3.3	65	2.7	3	10	5	N/A
[78]	2.45	26.3/33	2	90	1.88	2	9	7	3
[79]	2.2	43	1.2	65	6.25	2	6.2	4	2.1
This work	1.9	35.3/44.9	2.5	40	4	4	13	7.36	9

VII.F. Conclusion

A 1.9 GHz segmented linear PA was designed and implemented in 40 nm CMOS technology. The input signal is segmented and strategically amplified in the digital domain, while the PA is segmented and its segments are properly manipulated to maintain its power gain invariant with voltage while achieving significant power savings. The architecture emulates the operation of the conventional class-B amplifier, thus achieving similar power efficiency. However, the fact that the PA drivers are made switchable, then this architecture may result in better power efficiency. The PA achieved a saturated/maximum linear output power of 35/31 dBm with corresponding peak PAEs of 44.9% and 28.8%, respectively. A fast yet efficient

switching scheme that employs direct coupling between PA sections and drivers was demonstrated, which enabled the PA to improve efficiency in the PBO region within a wideband communication standard. The architecture can be combined with envelope tracking techniques to achieve better power efficiency figures. The proposed techniques are general and can be used in other PA architectures as well.

VIII. CONCLUSION

This dissertation has examined various important aspects related to CMOS RF PA design, particularly in the context of wireless communication applications. The proposals, and theoretical analyses in early sections have led to the design and implementation of a high efficiency and high performance PA in 40 nm CMOS technology. The segmented PA is able to switch between different modes within a very short time frame, and thus has achieved very high average power efficiency while maintain an industrial-level good linearity. The fast switching feature is the first to be reported in the RF PA research and development community. This work demonstrates that CMOS technology can be a serious candidate for implementing high power RF PA, among other more expensive technologies such as SiGe BiCMOS and GaAs MESFET.

Section II first introduces various operation modes of PAs, which serve as a starting point of the discussion and analysis of RF PA design. While it turned out the work in this dissertation is mainly focused on linear PA design, the design and simulation techniques, especially the segmentation approach, can be employed in switching mode PAs too. Also, the detailed analysis of the operation of PA in different modes in this section serves as the basis on which the switching of segmented PA architecture forms.

The introduction of some of the relevant concepts in digital communication theories in Section III points out the importance of linear PA in the current and future generations of wireless communications. Therefore, it is indispensable to study the nonlinear mechanism in RF PA in Section IV. The root cause of nonlinear effects lies in the use of nonlinear active device, and MOSFET nonlinear mechanisms, including channel-length modulation, velocity saturation, and mobility degradation, are analyzed in details. It is also very important to have a good understanding of how the nonlinear effects can be efficiently characterized in design. Conventional linearity tests, i.e. single-tone and two-tone tests are also briefly reviewed. As mentioned in Section III, modern communications use wideband, multicarrier modulation schemes, hence the conventional linearity tests can only provide an indication of the PA linearity. On the other hand, a multitone test or modulated envelope simulation require long simulation time and large computation resource. To resolve this issue, a detailed analysis of multitone nonlinear effects is carried out in this section, and as the analysis shows, a simple two-tone test can be used to predict multitone nonlinear behavior. The result of this innovative analysis could save much resource in the design and implementation of RF PA. Although currently not a major nonlinearity contributor to CMOS PA, the amplitude modulation to phase modulation (AM-to-PM) conversion is briefly analyzed at the end of this section.

The analysis and design of impedance matching network is discussed in Section V. This topic is in some way unique in RF design, and usually sets apart RF design from analog design. Yet at least for RF

PA applications, impedance matching is very important. A mismatched impedance at the output of the PA would not only degrade the power efficiency by a lot, but may even damage the device due to the high power levels. Therefore, in this section the many design considerations of impedance matching circuits are analyzed, including bandwidth, harmonic rejection, and insertion loss. Basic impedance matching building blocks are also analyzed, with an emphasis in the Π -match due to its versatility and generality.

A theoretical study of nearly all PA architectures is presented in Section VI, which shows that, despite the huge amount of effort of research and development, the PA architectures still usually improve the power efficiency *or* linearity, but not both. This is because this two aspects is the fundamental tradeoff in PA design.

Finally, based on all the theoretical studies and proposed innovations on the previous sections, a segmented PA combined with digital pre-warp architecture is proposed, designed, and implemented in 40 nm CMOS technology. This work reconciled the fundamental tradeoff between power efficiency and linearity, and the fast-switching scheme implemented in this system has resulted in switching of PA segments within the modulation, which is the first to be reported. The PA achieved 35 dBm output power and 38 dB power gain in the 1.9 GHz WCDMA band. The peak PAE of 44.9% and good linearity meeting 3GPP standard shows that this PA has a very balanced tradeoff between efficiency and linearity, and is competitive with other more expensive, traditional RF IC technologies, such as SiGe, GaAs, and InP.

REFERENCES

- [1] C. Paul, "Telephones aboard the Metroliner," *Bell Laboratories Record*, March 1969.
- [2] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed. New York, NY, USA: Cambridge University Press, 2004.
- [3] International Telecommunication Union, *International Telecommunication Union Database*, 2015. [Online]. Available: www.itu.int
- [4] World Bank, *World Bank Database*, 2016. [Online]. Available: www.worldbank.org
- [5] International Data Corporation, *International Data Corporation Database*, 2016. [Online]. Available: www.idc.com
- [6] J. Crols and M. Steyaert, "A single-chip 900 MHz CMOS receiver front-end with a high performance low-IF topology," *Solid-State Circuits, IEEE Journal of*, vol. 30, no. 12, pp. 1483–1492, Dec 1995.
- [7] A. Rofougaran, J.-C. Chang, M. Rofougaran, and A. Abidi, "A 1 GHz CMOS RF front-end IC for a direct-conversion wireless receiver," *Solid-State Circuits, IEEE Journal of*, vol. 31, no. 7, pp. 880–889, Jul 1996.
- [8] J. Rudell, J.-J. Ou, T. Cho, G. Chien, F. Brianti, J. Weldon, and P. Gray, "A 1.9-GHz wide-band IF double conversion CMOS receiver for cordless telephone applications," *Solid-State Circuits, IEEE Journal of*, vol. 32, no. 12, pp. 2071–2088, Dec 1997.
- [9] A. Shahani, D. Shaeffer, and T. Lee, "A 12-mW wide dynamic range CMOS front-end for a portable GPS receiver," *Solid-State Circuits, IEEE Journal of*, vol. 32, no. 12, pp. 2061–2070, Dec 1997.
- [10] R. Kulkarni, J. Kim, H.-J. Jeon, J. Xiao, and J. Silva-Martinez, "UHF receiver front-end: Implementation and analog baseband design considerations," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 20, no. 2, pp. 197–210, Feb 2012.
- [11] O. Erdogan, R. Gupta, D. Yee, J. Rudell, J.-S. Ko, R. Brockenbrough, S.-O. Lee, E. Lei, J. L. Tham, H. Wu, C. Conroy, and B. Kim, "A single-chip quad-band GSM/GPRS transceiver in 0.18 μm standard CMOS," in *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International*, Feb 2005, pp. 318–601 Vol. 1.
- [12] D. Kaczman, M. Shah, N. Godambe, M. Alam, H. Guimaraes, L. Han, M. Rachedine, D. Cashen, W. Getka, C. Dozier, W. Shepherd, and K. Coughlar, "A single-chip tri-band (2100, 1900, 850/800 MHz) WCDMA/HSDPA cellular transceiver," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 5, pp. 1122–1132, May 2006.

- [13] A. Hadjichristos, M. Cassia, H. Kim, C. H. Park, K. Wang, W. Zhuo, B. Ahrari, R. Brockenbrough, J. Chen, C. Donovan, R. Jonnalagedda, J. Kim, J. Ko, H. Lee, S. Lee, E. Lei, T. Nguyen, T. Pan, S. Sridhara, W. Su, H. Yan, J. Yang, C. Conroy, C. Persico, K. Sahota, and B. Kim, "Single-chip RF CMOS UMTS/EGSM transceiver with integrated receive diversity and GPS," in *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, Feb 2009, pp. 118–119, 119a.
- [14] H. Moon, J. Han, S.-I. Choi, D. Keum, and B.-H. Park, "An area-efficient 0.13- μ m CMOS multiband WCDMA/HSDPA receiver," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 58, no. 5, pp. 1447–1455, May 2010.
- [15] H. Wang, C.-H. Peng, C. Lu, Y. Chang, R. Huang, A. Chang, G. Shih, R. Hsu, P. Liang, S. Son, A. Niknejad, G. Chien, C. Tsai, and H. Hwang, "A highly-efficient multi-band multi-mode digital quadrature transmitter with 2D pre-distortion," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, May 2013, pp. 501–504.
- [16] Anonymous, "The great debate: SOC vs. SIP," *EE Times*, March 2005.
- [17] I. Aoki, S. Kee, R. Magoon, R. Aparicio, F. Bohn, J. Zachan, G. Hatcher, D. McClymont, and A. Hajimiri, "A fully-integrated quad-band gsm/gprs cmos power amplifier," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 12, pp. 2747–2758, Dec 2008.
- [18] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2007.
- [19] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Potheary, J. Sevic, and N. Sokal, "Power amplifiers and transmitters for RF and microwave," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 50, no. 3, pp. 814–826, Mar 2002.
- [20] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed. Norwood, MA, USA: Artech House, Inc., 2006.
- [21] S.-A. El-Hamamsy, "Design of high-efficiency RF class-D power amplifier," *Power Electronics, IEEE Transactions on*, vol. 9, no. 3, pp. 297–308, May 1994.
- [22] N. Sokal and A. Sokal, "Class E-a new class of high-efficiency tuned single-ended switching power amplifiers," *Solid-State Circuits, IEEE Journal of*, vol. 10, no. 3, pp. 168–176, Jun 1975.
- [23] F. Raab, "Idealized operation of the class E tuned power amplifier," *Circuits and Systems, IEEE Transactions on*, vol. 24, no. 12, pp. 725–735, Dec 1977.
- [24] —, "Class-F power amplifiers with maximally flat waveforms," *Microwave Theory and Techniques*,

IEEE Transactions on, vol. 45, no. 11, pp. 2007–2012, Nov 1997.

- [25] A. Hadjichristos, “Transmit architectures and power control schemes for low cost highly integrated transceivers for GSM/EDGE applications,” in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, vol. 3, May 2003, pp. III–610–III–613 vol.3.
- [26] H. Packard, “Digital modulation in communications systems - an introduction,” *Hewlett Packard Application Note 1298*, July 1997.
- [27] H. Izumi, M. Kojima, Y. Umeda, and O. Takyu, “Comparison between quadrature- and polar-modulation switching-mode transmitter with pulse-density modulation,” in *Advanced Communication Technology (ICACT), 2013 15th International Conference on*, Jan 2013, pp. 1140–1145.
- [28] B. Razavi, *RF Microelectronics*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2011.
- [29] 3rd Generation Partnership Project, “3gpp ts 05.05 technical specification rev. 8.20.0,” 1999. [Online]. Available: <http://www.3gpp.org>
- [30] I. S. Committee, “Ieee std. 802.11ac-2013,” *IEEE Standard for Information Technology*, pp. 1–425, Dec 2013.
- [31] 3rd Generation Partnership Project, “3gpp ts 25.101 technical specification rev.12.3.0,” March 2014. [Online]. Available: <http://www.3gpp.org>
- [32] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2009.
- [33] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. New York, NY, USA: Oxford University Press, 1999.
- [34] S. Narayanan, “Transistor distortion analysis using Volterra series representation,” *Bell System Technical Journal*, The, vol. 46, no. 5, pp. 991–1024, May 1967.
- [35] S. Maas, “Volterra analysis of spectral regrowth,” *Microwave and Guided Wave Letters, IEEE*, vol. 7, no. 7, pp. 192–193, Jul 1997.
- [36] Q. Wu, H. Xiao, and F. Li, “Linear rf power amplifier design for cdma signals: a spectrum analysis approach,” *Microwave Journal*, vol. 41, no. 12, p. 22, 1998.
- [37] J. Pedro and N. de Carvalho, “On the use of multitone techniques for assessing rf components’ intermodulation distortion,” *Microwave Theory and Techniques, IEEE Transactions on*, vol. 47, no. 12, pp. 2393–2402, Dec 1999.
- [38] S. A. Maas, *Nonlinear Microwave and RF Circuits*, 2nd ed. Norwood, MA, USA: Artech House, Inc., 2002.

- [39] K. Kundert, "Introduction to RF simulation and its application," *Solid-State Circuits, IEEE Journal of*, vol. 34, no. 9, pp. 1298–1319, Sep 1999.
- [40] T. Quarles, D. Pederson, R. Newton, A. Sangiovanni-Vincentelli, and C. Wayne. SPICE User Guide. EECS Department of the University of California at Berkeley. [Online]. Available: <http://bwrcs.eecs.berkeley.edu/Classes/IcBook/SPICE/>
- [41] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th ed. New York, NY, USA: Oxford University Press, 2008.
- [42] M. Leffel, "Intermodulation distortion in a multi-signal environment," *RF Design*, June 1995.
- [43] H. Qian and J. Silva-Martinez, "A 44.9% PAE digitally-assisted linear power amplifier in 40 nm CMOS," in *Solid-State Circuits Conference (A-SSCC), 2014 IEEE Asian*, Nov 2014, pp. 349–352.
- [44] J. Aikio and T. Rahkonen, "A comprehensive analysis of AM-AM and AM-PM conversion in an LDMOS RF power amplifier," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 57, no. 2, pp. 262–270, Feb 2009.
- [45] L. Cotimos Nunes, P. Cabral, and J. Pedro, "AM/AM and AM/PM distortion generation mechanisms in Si LDMOS and GaN HEMT based RF power amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 62, no. 4, pp. 799–809, April 2014.
- [46] G. Gonzalez, *Microwave Transistor Amplifiers : Analysis and Design*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.
- [47] S. Cripps, "A theory for the prediction of GaAs FET load-pull power contours," in *Microwave Symposium Digest, 1983 IEEE MTT-S International*, May 1983, pp. 221–223.
- [48] G. E. Bodway, "Two port power flow analysis using generalized scattering parameters," *Microwave Journal*, vol. 10, no. 6, May 1967, also available in HP application note 95.
- [49] J. Wannstrom, *Carrier Aggregation Explained*, June 2013. [Online]. Available: www.3gpp.org
- [50] Qualcomm Technology Inc., "LTE advanced - evolving and expanding in to new frontiers," August 2014. [Online]. Available: www.qualcomm.com
- [51] H. Chireix, "High power outphasing modulation," *Radio Engineers, Proceedings of the Institute of*, vol. 23, no. 11, pp. 1370–1392, Nov 1935.
- [52] I. Hakala, D. Choi, L. Gharavi, N. Kajakine, J. Koskela, and R. Kaunisto, "A 2.14-GHz Chireix outphasing transmitter," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 53, no. 6, pp. 2129–2138, June 2005.
- [53] F. Raab, "Efficiency of outphasing RF power-amplifier systems," *Communications, IEEE Transactions*

on, vol. 33, no. 10, pp. 1094–1099, Oct 1985.

- [54] W. Doherty, “A new high-efficiency power amplifier for modulated waves,” *Bell System Technical Journal, The*, vol. 15, no. 3, pp. 469–475, July 1936.
- [55] F. Raab, “Efficiency of Doherty RF power-amplifier systems,” *Broadcasting, IEEE Transactions on*, vol. BC-33, no. 3, pp. 77–83, Sept 1987.
- [56] L. Kahn, “Single-sideband transmission by envelope elimination and restoration,” *Proceedings of the IRE*, vol. 40, no. 7, pp. 803–806, July 1952.
- [57] J. Staudinger, B. Gilsdorf, D. Newman, G. Norris, G. Sadowiczak, R. Sherman, and T. Quach, “High efficiency CDMA RF power amplifier using dynamic envelope tracking technique,” in *Microwave Symposium Digest. 2000 IEEE MTT-S International*, vol. 2, June 2000, pp. 873–876 vol.2.
- [58] G. Hanington, P.-F. Chen, P. Asbeck, and L. Larson, “High-efficiency power amplifier using dynamic power-supply voltage for CDMA applications,” *Microwave Theory and Techniques, IEEE Transactions on*, vol. 47, no. 8, pp. 1471–1476, Aug 1999.
- [59] F. Wang, D. Kimball, D. Lie, P. Asbeck, and L. Larson, “A monolithic high-efficiency 2.4-GHz 20-dBm SiGe BiCMOS envelope-tracking OFDM power amplifier,” *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 6, pp. 1271–1281, June 2007.
- [60] J. Kim, Y. Yoon, H. Kim, K. H. An, W. Kim, H.-W. Kim, C.-H. Lee, and K. Kornegay, “A linear multi-mode CMOS power amplifier with discrete resizing and concurrent power combining structure,” *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 5, pp. 1034–1048, May 2011.
- [61] A. Shirvani, D. Su, and B. Wooley, “A CMOS RF power amplifier with parallel amplification for efficient power control,” *Solid-State Circuits, IEEE Journal of*, vol. 37, no. 6, pp. 684–693, Jun 2002.
- [62] P. Reynaert and M. S. Steyaert, “A 2.45-GHz 0.13- μm CMOS PA with parallel amplification,” *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 3, pp. 551–562, March 2007.
- [63] G. Liu, P. Haldi, T.-J. K. Liu, and A. Niknejad, “Fully integrated CMOS power amplifier with efficiency enhancement at power back-off,” *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 3, pp. 600–609, March 2008.
- [64] Y. Yoon, J. Kim, H. Kim, K. H. An, O. Lee, C.-H. Lee, and J. Kenney, “A dual-mode CMOS RF power amplifier with integrated tunable matching network,” *Microwave Theory and Techniques, IEEE Transactions on*, vol. 60, no. 1, pp. 77–88, Jan 2012.
- [65] A. Niknejad, D. Chowdhury, and J. Chen, “Design of CMOS power amplifiers,” *Microwave Theory and Techniques, IEEE Transactions on*, vol. 60, no. 6, pp. 1784–1796, June 2012.

- [66] L. Larson, "RF and microwave hardware challenges for future radio spectrum access," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 321–333, March 2014.
- [67] K. Okada, R. Minami, Y. Tsukui, S. Kawai, Y. Seo, S. Sato, S. Kondo, T. Ueno, Y. Takeuchi, T. Yamaguchi, A. Musa, R. Wu, M. Miyahara, and A. Matsuzawa, "A 64-QAM 60GHz CMOS transceiver with 4-channel bonding," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, Feb 2014, pp. 346–347.
- [68] M. Ebrahimi, M. Helaoui, and F. Ghannouchi, "Delta-sigma-based transmitters: Advantages and disadvantages," *Microwave Magazine, IEEE*, vol. 14, no. 1, pp. 68–78, Jan 2013.
- [69] E. Kaymaksut and P. Reynaert, "A dual-mode transformer-based doherty LTE power amplifier in 40nm CMOS," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, Feb 2014, pp. 64–65.
- [70] W.-Y. Kim, H. Son, J. Kim, J. Jang, I. Oh, and C. Park, "A CMOS envelope-tracking transmitter with an on-chip common-gate voltage modulation linearizer," *Microwave and Wireless Components Letters, IEEE*, vol. PP, no. 99, pp. 1–1, 2014.
- [71] K. Oishi, E. Yoshida, Y. Sakai, H. Takauchi, Y. Kawano, N. Shirai, H. Kano, M. Kudo, T. Murakami, T. Tamura, S. Kawai, S. Yamaura, K. Suto, H. Yamazaki, and T. Mori, "A 1.95GHz fully integrated envelope elimination and restoration CMOS power amplifier with envelope/phase generator and timing aligner for WCDMA and LTE," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, Feb 2014, pp. 60–61.
- [72] B. Sahu and G. Rincon-Mora, "A high-efficiency linear RF power amplifier with a power-tracking dynamically adaptive buck-boost supply," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 52, no. 1, pp. 112–120, Jan 2004.
- [73] I. Rippke, J. Duster, and K. Kornegay, "A single-chip variable supply voltage power amplifier," in *Radio Frequency integrated Circuits (RFIC) Symposium, 2005. Digest of Papers. 2005 IEEE*, June 2005, pp. 255–258.
- [74] D. Chowdhury, C. Hull, O. Degani, Y. Wang, and A. Niknejad, "A fully integrated dual-mode highly linear 2.4 GHz CMOS power amplifier for 4G WiMax applications," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 12, pp. 3393–3402, Dec 2009.
- [75] H. Hedayati, M. Mobarak, G. Varin, P. Meunier, P. Gamand, E. Sanchez-Sinencio, and K. Entesari, "A 2-GHz highly linear efficient dual-mode BiCMOS power amplifier using a reconfigurable matching network," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 10, pp. 2385–2404, Oct 2012.

- [76] A. Afsahi and L. Larson, "Monolithic power-combining techniques for watt-level 2.4-GHz CMOS power amplifiers for WLAN applications," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 61, no. 3, pp. 1247–1260, March 2013.
- [77] P. T. M. van Zeijl and M. Collados, "A digital envelope modulator for a wlan ofdm polar transmitter in 90 nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 10, pp. 2204–2211, Oct 2007.
- [78] E. Kaymaksut and P. Reynaert, "Transformer-based uneven doherty power amplifier in 90 nm cmos for wlan applications," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 7, pp. 1659–1671, July 2012.
- [79] L. Ye, J. Chen, L. Kong, P. Cathelin, E. Alon, and A. Niknejad, "A digitally modulated 2.4GHz WLAN transmitter with integrated phase path and dynamic load modulation in 65nm CMOS," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, Feb 2013, pp. 330–331.